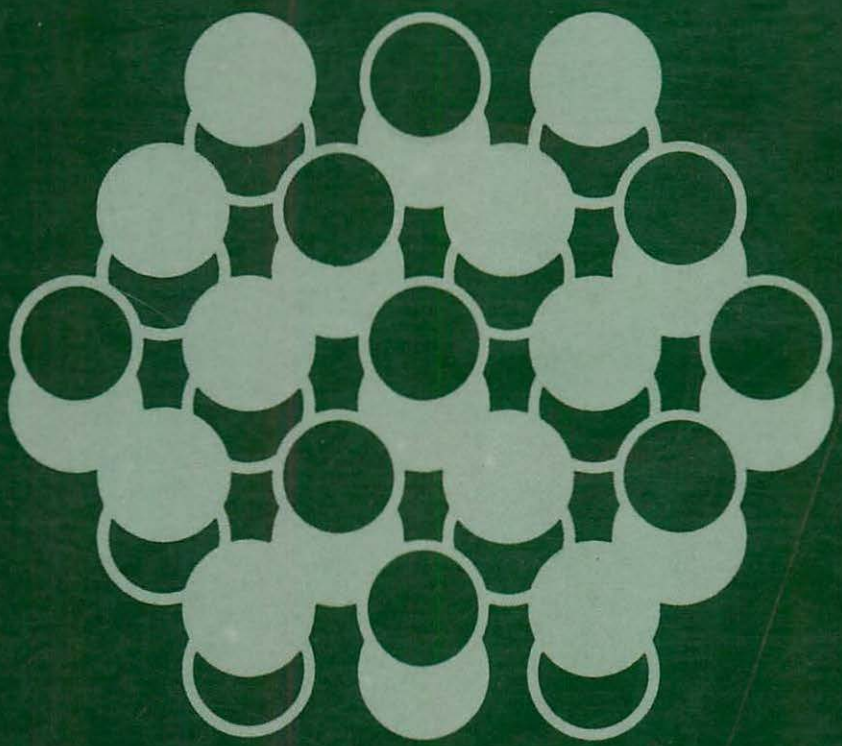


TEXTBOOK OF BIOCHEMISTRY WITH CLINICAL CORRELATIONS

Thomas M. Devlin, EDITOR

RICHARD PEARSON
M2 Gp 16
285 BROADSTONE Rd.
HEATON CHAPEL,
STOCKPORT
432 2469

21·00

TEXTBOOK OF BIOCHEMISTRY

TEXTBOOK OF
with Clinical

BIOCHEMISTRY Correlations

Edited by

Thomas M. Devlin, Ph.D.

Professor and Chairman
Department of Biological Chemistry
Hahnemann Medical College and Hospital
Philadelphia, Pennsylvania

A WILEY MEDICAL PUBLICATION
JOHN WILEY & SONS / New York • Chichester • Brisbane • Toronto • Singapore

Cover and interior design: Wanda Lubelska
Production Editor: David M. Feren

Library of Congress Cataloging in Publication Data
Main entry under title:

Textbook of biochemistry.

(A Wiley medical publication)
Includes index.
1. Biological chemistry. 2. Physiology, Pathological. I. Devlin, Thomas M. II. Series.
QP514.2.T4 612'.015 81-13063
AACR2

Printed in Singapore

10 9 8 7 6 5

To Marjorie

Contributors

Stelios Aktipis, Ph.D.
Professor
Department of Biochemistry
Loyola University Stritch School of Medicine
Maywood, Illinois

William Awad, Jr., M.D., Ph.D.
Professor
Departments of Medicine and Biochemistry
University of Miami School of Medicine
Miami, Florida

James Baggott, Ph.D.
Associate Professor
Department of Biological Chemistry
Hahnemann Medical College and Hospital
Philadelphia, Pennsylvania

Stephen G. Chaney, Ph.D.
Assistant Professor
Department of Biochemistry and Nutrition
School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

Joseph G. Cory, Ph.D.
Professor and Chairman
Department of Biochemistry
College of Medicine
University of South Florida
Tampa, Florida

David W. Crabb, M.D.
Fellow, Gastroenterology
Department of Medicine
Indiana University School of Medicine
Indianapolis, Indiana

Thomas M. Devlin, Ph.D.
Professor and Chairman
Department of Biological Chemistry
Hahnemann Medical College and Hospital
Philadelphia, Pennsylvania

Thomas I. Diamondstone, Ph.D.
Assistant Professor
Department of Physiology and Biochemistry
The Medical College of Pennsylvania and Hospital
Philadelphia, Pennsylvania

Edward Glassman, Ph.D.
Professor
Department of Biochemistry and Nutrition
School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

Robert H. Glew, Ph.D.
Associate Professor
Department of Biochemistry
School of Medicine
University of Pittsburgh
Pittsburgh, Pennsylvania

Peter N. Gray, Ph.D.
Associate Professor
Department of Biochemistry and Molecular Biology
University of Oklahoma Health Sciences Center
Oklahoma City, Oklahoma

Robert A. Harris, Ph.D.
Professor
Department of Biochemistry
Indiana University School of Medicine
Indianapolis, Indiana

Ulrich Hopfer, M.D., Ph.D.
Associate Professor
Department of Anatomy
School of Medicine
Case Western Reserve University
Cleveland, Ohio

Francis N. LeBaron, Ph.D.
Professor
Department of Biochemistry
School of Medicine
University of New Mexico
Albuquerque, New Mexico

Karl H. Muench, M.D.
Professor
Departments of Medicine and Biochemistry
University of Miami School of Medicine
Miami, Florida

Merle S. Olson, Ph.D.
Professor
Department of Biochemistry
University of Texas Health Center at San Antonio
San Antonio, Texas

Richard M. Schultz, Ph.D.
Associate Professor
Department of Biochemistry and Biophysics
Loyola University Stritch School of Medicine
Maywood, Illinois

Nancy B. Schwartz, Ph.D.
Associate Professor
Departments of Pediatrics and Biochemistry
University of Chicago
Chicago, Illinois

Frank Ungar, Ph.D.
Professor
Department of Biochemistry
Medical School
University of Minnesota
Minneapolis, Minnesota

John F. Van Pilsum, Ph.D.
Professor
Department of Biochemistry
Medical School
University of Minnesota
Minneapolis, Minnesota

Marilyn S. Wells, M.D.
Adjunct Assistant Professor
Department of Medicine
University of Miami School of Medicine
Chief, Spinal Cord Injury Service
Veterans Administration Medical Center
Miami, Florida

J. Lyndal York, Ph.D.
Professor
Department of Biochemistry
School of Medicine
University of Arkansas
Little Rock, Arkansas

Preface

This textbook deals with the biochemistry of mammalian cells and relates the biochemical events at the cellular level to the physiological processes occurring in the whole animal. The topics were selected to meet the needs of a medical school course in biochemistry, and, wherever possible, information or examples from our knowledge of the biochemistry of humans is presented. The content is, in addition, applicable to a first course for upper-level undergraduate or graduate students.

The first six chapters cover the areas of protein and enzyme chemistry, membrane structure, and bioenergetics, setting the stage for discussions of the metabolism of the principal cellular components. Presentation of metabolic interrelationships at the cellular and tissue level is followed by chapters on the major hormones. The biochemical events in the transmittal of genetic information and the mechanism of phenotypic expression and its control, as well as recent developments in recombinant DNA research and genetic engineering, are discussed. The remaining chapters cover essential aspects of physiological chemistry, including the biochemistry of selected mammalian tissues, iron and heme metabolism, gas transport, regulation of pH, and digestion and absorption of foodstuffs. The textbook concludes with a discussion of nutrition from a biochemical perspective.

In order to emphasize the relevancy of the topics to disease problems, each chapter includes selected clinical correlations relating normal cellular biochemical events to pathophysiological states in humans. The correlations are intended to describe the aberrant biochemistry of the disease state rather than specific case reports. In some instances the same clinical condition is presented in different chapters from a different perspective. The clinical correlations are presented as separate entities in order not to interrupt the flow of the biochemistry discussion. All pertinent biochemical information is presented in the main text, and an understanding of the material does not require a detailed review of the correlations. In some chapters, however,

clinical discussions are part of the principal text because of the close relationship of some topics to clinical conditions.

As with any textbook, selection of material to be presented was a difficult problem. Much of our knowledge of intracellular chemical events is derived from studies of single cells, such as bacteria and yeast; in fact, for some topics we have very sketchy knowledge about the events in mammalian cells. Thus, in some sections it has been necessary to discuss information derived from bacterial systems as a model for what may occur in the more complex mammalian tissue.

The individual contributors were requested to prepare their chapters for a teaching textbook. The work is not simply a compendium of biochemical facts. We have attempted to prevent the narrative from becoming too encyclopedic in depth but still to contain sufficient detail to make the book useful as a resource for biochemical concepts and information. Neither is the book intended as a review of the available literature. Each author was requested not to refer to specific researchers. Our apologies to those many biochemists who rightfully should be acknowledged for their outstanding research contributions to the field of biochemistry. Each chapter contains a bibliography that can be used as a guide to the primary literature in the field.

A textbook with many contributors, each writing in his own area of expertise, has many positive aspects, but also some potential problems. Before a commitment to this project was made, the pros and cons of such an approach were carefully evaluated. The deciding factor was the value of having a number of scientist–teachers contribute sections in their area of biochemical interest. Each author has been involved actively in teaching biochemistry in a medical school and has an active research interest in the field in which he has written. Thus, each has the perspective of the classroom instructor and the experience to select the topics and determine the emphasis required for students in a course of biochemistry. Every contributor, however, brings to the book an individual writing style, leading to differences in presentation from chapter to chapter. It was decided, however, that this should not be an impediment to the student's ability to understand the material. Students are accustomed to learning from a variety of sources. As editor, I assumed the responsibility for keeping to a minimum redundancies and inconsistencies in content. Some repetitions of topics in different chapters were retained because they were considered of such importance that reiteration would be helpful to the reader.

In any project, one person must accept the responsibility for the final product. The decisions concerning selection of topics and format, reviewing the drafts, and responsibility for the final checking of the book were entirely mine. I would welcome comments, criticisms, and suggestions from the faculty and students who use this textbook. It is our hope that this work will be of value to those embarking on the exciting experience of learning biochemistry.

Thomas M. Devlin

Acknowledgments

This project would never have come to fruition without the encouragment and participation of many individuals. Each of the contributors received the support of associates and students in the preparation of his chapter, and, for fear of omitting someone, it was decided not to acknowledge individuals by name. To everyone who gave of their time unselfishly and shared their objective and critical evaluations of the text, we extend our sincerest thanks. In addition, every contributor has been influenced by former teachers and colleagues, various reference resources, and, of course, the research literature of biochemistry; we are deeply indebted to these many sources of inspiration.

As editor, I extend a very special thanks to all of the contributors for accepting the challenge of preparing the chapters, for sharing ideas and making recommendations to improve the book, for accepting so readily suggestions to modify their contributions, and for their cooperation throughout the period of preparation. My personal and deep appreciation goes to the staff of John Wiley & Sons, Inc., who worked with me during these many months. I am indebted to Mr. John de Carville, Executive Editor, whose foresight and patience assisted me in the early stages of planning and who gave me his unqualified support during the preparation and review of the manuscripts. Special recognition is extended to Mr. Charles Kyreakou, Managing Editor, for his able assistance in the initial period of production; to Mr. Bruce Williams, Editorial Supervisor, for carefully reviewing the manuscripts; to Ms. Deslie Lawrence, who ably copy edited the manuscript; and to Mr. David Feren, Assistant Production Manager, who led me through the intricacies of production with a steady and patient hand. A special thanks is due the Illustration Department of John Wiley, particularly Mr. Ishaya Monokoff, Illustration Production Manager, and the many staff members who assisted in the completion of the book.

I am deeply indebted to my staff during the last few years, including Mrs. Caren Gold, Ms. Mary Ellen Minto, Mrs. Eileen O'Shaughnessy, and especially Mrs.

Mary Hart vanBoven, all of whom faithfully and efficiently completed the multitude of little chores involved in this project. To each I extend my heartfelt appreciation.

Finally, a very special thanks to a supportive and considerate family, and particularly to my wife, Marjorie, who had the foresight to encourage me to undertake this project and who created an environment in which I could devote the many hours required to bring together the elements of this textbook. To all, my deepest and sincerest thanks.

Thomas M. Devlin

Contents

ENZYMES: CLASSIFICATION, KINETICS, AND CONTROL 137

J. Lyndal York

CLINICAL CORRELATIONS

5

BIOLOGICAL MEMBRANES: STRUCTURE AND MEMBRANE TRANSPORT 217

Thomas M Devlin

CLINICAL CORRELATIONS

BIOENERGETICS AND OXIDATIVE METABOLISM 255

Merle S. Olson

CLINICAL CORRELATIONS

7

CARBOHYDRATE METABOLISM I: MAJOR METABOLIC PATHWAYS AND THEIR CONTROL 325

Robert A. Harris

CLINICAL CORRELATIONS

CARBOHYDRATE METABOLISM II: SPECIAL PATHWAYS 405

Nancy B. Schwartz

CLINICAL CORRELATIONS

12

AMINO ACID METABOLISM II: METABOLISM OF THE INDIVIDUAL AMINO ACIDS 563

Thomas I. Diamondstone

CLINICAL CORRELATIONS

13

PURINE AND PYRIMIDINE NUCLEOTIDE METABOLISM 627

Joseph G Cory

CLINICAL CORRELATIONS

14

METABOLIC INTERRELATIONSHIPS 681

Robert A. Harris and David W. Crabb

CLINICAL CORRELATIONS

RNA: TRANSCRIPTION AND POSTTRANSCRIPTIONAL MODIFICATION 871

Peter N. Gray

CLINICAL CORRELATIONS

PROTEIN SYNTHESIS: TRANSLATION AND POSTTRANSLATIONAL MODIFICATION 917

Karl H. Muench

CLINICAL CORRELATIONS

20

REGULATION OF GENE EXPRESSION 951

Edward Glassman

CLINICAL CORRELATIONS

21

METABOLISM OF INDIVIDUAL TISSUES 1001

John F. Van Pilsum

CLINICAL CORRELATIONS

22

IRON AND HEME METABOLISM 1061

William M. Awad, Jr. and Marilyn S. Wells

CLINICAL CORRELATIONS

23

GAS TRANSPORT AND pH REGULATION 1087

James Baggott

CLINICAL CORRELATIONS

CLINICAL CORRELATIONS

PRINCIPLES OF NUTRITION II: MICRONUTRIENTS 1197

Stephen G. Chaney

CLINICAL CORRELATIONS

1

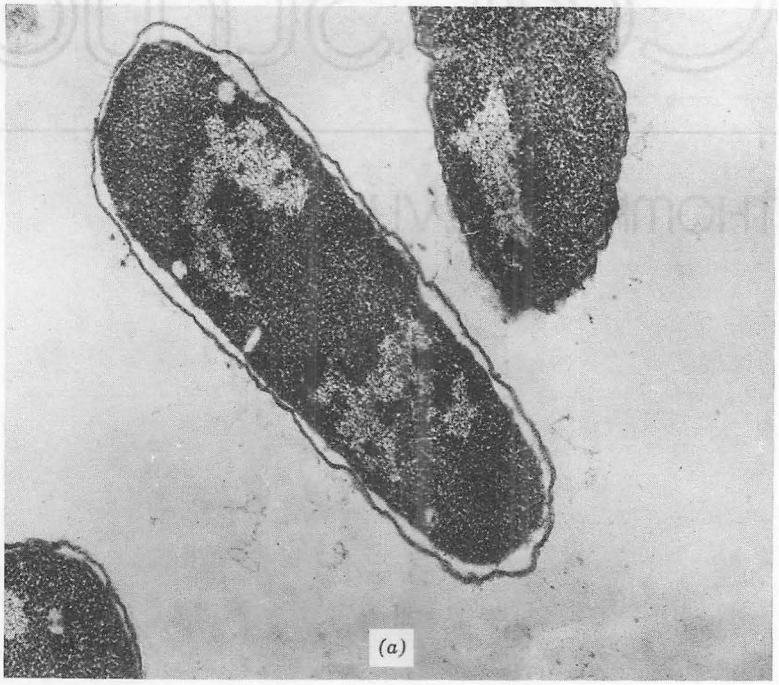
Eucaryotic Cell Structure

THOMAS M. DEVLIN

1.1 CELLS AND CELLULAR COMPARTMENTS

By a process not yet understood and in a time span few comprehend, elements such as carbon, oxygen, hydrogen, nitrogen, phosphorus, and sulfur combined, dispersed, and recombined to form a variety of compounds, until that unique combination came together with the required information and capability to reproduce itself. Life as we know it had occurred. With life occurred the establishment of a unit of space with a definable environment surrounded by a membrane, that is, a cell. With the passing of time a diversity of cells evolved and their chemistry and structure became more complex. Many of the chemical reactions occurring in cells can be duplicated in test tubes, but the challenge of biochemical research is to unravel the mechanisms behind the organized and controlled manner in which cells carry out these reactions.

One of the most important structures of all cells is the limiting outer membrane, termed the plasma membrane, which delineates the intracellular and extracellular environment. The plasma mem-



brane separates the variable and potentially hostile environment outside the cell from the relatively constant milieu within the cell and is the communication link between the cell and its surroundings. It is the plasma membrane that delineates the space occupied by a cell.

Based on both microscopic and biochemical differences, living cells are divided into two major classes, the procaryotes and the eucaryotes. Procaryotic cells, which include bacteria, blue-green algae, and rickettsia, lack extensive intracellular anatomy (Figure 1.1). Intracellular structures of cells are due to the presence of macromolecules or membrane systems, which can be visualized in a microscope under appropriate conditions. The deoxyribonucleic

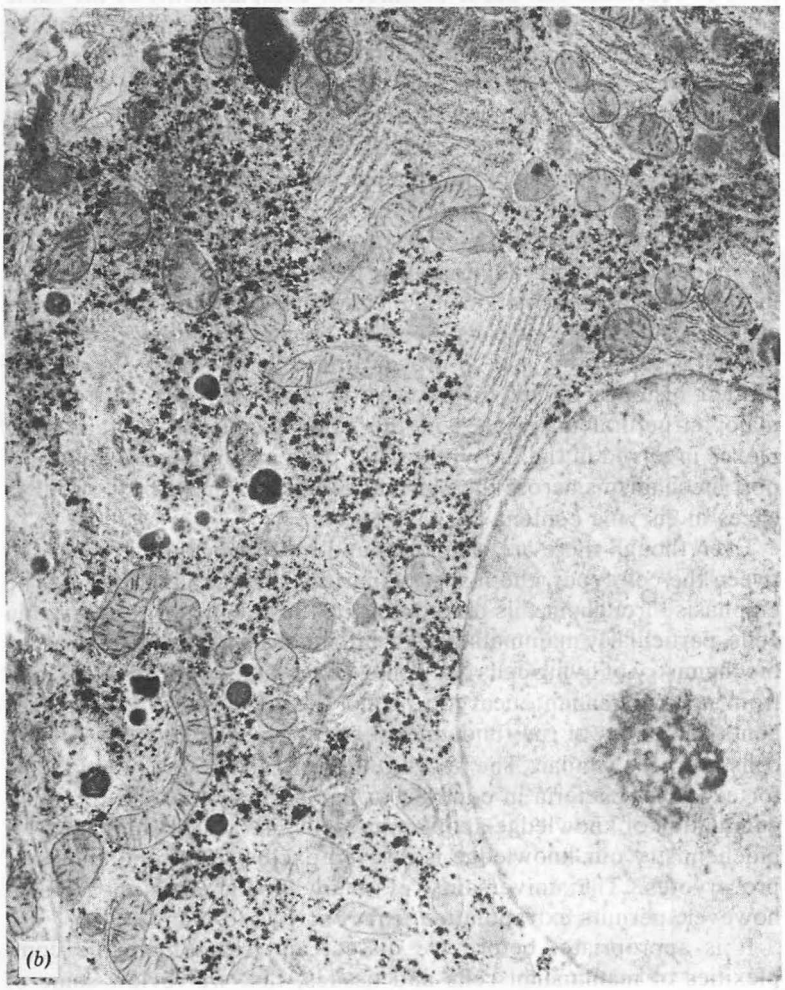


Figure 1.1
The cellular organization of procaryatic and eucaryotic cells.
(*a*) *An electron micrograph of an* E. coli, *a representative procaryote; approximate magnification 30,000×. There is little apparent intracellular organization and no cytoplasmic organelles. The chromatin is condensed into a nuclear zone but is not surrounded by a membrane. Procaryotic cells are smaller than eucaryotic cells.*

Photograph generously supplied by Dr. M. E. Bayer.

(*b*) *An electron micrograph of a thin section of a rat liver cell (hepatocyte), a representative eucaryotic cell; approximate magnification 7500×. Note the distinct nuclear membrane, the variety of membrane bound organelles or vesicles, and the extensive membrane systems. The various membranes create a variety of intracellular compartments.*

Photograph reprinted with permission of Dr. K. R. Porter from Porter, K. R., and Bonneville, M. A. *Fine Structure of Cells and Tissues,* Philadelphia: Lea and Febiger, 1972.

acid (DNA) of procaryotes is often segregated into discrete masses, but it is not surrounded by a membrane or envelope. The plasma membrane is often observed to be invaginated, but there are no definable subcellular organized bodies in procaryotes.

In contrast to the procaryotes, eucaryotic cells, which include yeast, fungi, plant, and animal cells, have a well-defined nuclear membrane and a variety of intracellular structures and organelles. The intracellular membrane system establishes a number of distinct subcellular compartments permitting a unique degree of subcellular specialization. By compartmentalization of the cell, different chemical reactions requiring different environments can occur simultaneously. As an example, reactions requiring different ionic conditions or reactants can occur in different compartments of the same cell because the membranes surrounding a cellular compartment can control the environment in the compartment by regulating the movement of substances in and out. In addition, membranes are lipid in nature, and substances more soluble in the hydrophobic lipid environment will dissolve into the membrane. In fact, many biochemical reactions occur in specific membranes of the cell. The extensive membrane systems of the eucaryotic cell create an additional environment that the cell can use for its diverse functions.

Besides the structural variations, there are significant differences in the chemical composition and biochemical activities between procaryotic and eucaryotic cells. Some of the major differences between the cell types are the lack in procaryotic cells of a class of proteins, termed histones, which in eucaryotic cells complex with DNA; major structural differences in the ribonucleic acid protein complexes involved in the biosynthesis of proteins; differences in transport mechanisms across the plasma membrane; and a host of differences in enzyme content.

Even though there are structural and biochemical differences between the cell types, the many similarities are equally striking. The emphasis throughout this textbook is on the chemistry of eucaryotic cells, particularly mammalian cells, but much of our knowledge of the biochemistry of living cells has come from studies of procaryotic and from nonmammalian eucaryotic cells. The basic chemical components and many of the fundamental chemical reactions of all living cells are very similar. The accessibility of certain cell populations, for example, bacteria in contrast to human liver, has led to an accumulation of knowledge about some cells; in fact, in some areas of biochemistry our knowledge is derived exclusively from studies of procaryotes. The universality of many biochemical phenomena, however, permits extrapolation from bacteria to man.

It is appropriate, before we dissect and reassemble the complexities of mammalian cells and tissues, that we review some of

the chemical and physical characteristics of the environment in which the various biochemical phenomena occur, for it is important to recognize the constraints placed by the environment on biochemical reactions.

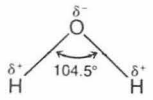


Figure 1.2
The structure of a water molecule.
The H—O—H bond angle is 104.5°, and both hydrogens carry a partial positive and the oxygen a partial negative charge, creating a dipole.

1.2 CELLULAR ENVIRONMENT–WATER AND SOLUTES

All biological cells depend on the environment for nutrients, which are requisites for their energy requirements, for replacement of components, and for growth. The composition of the external environment can vary significantly and cells have a variety of mechanisms to cope with these variations. In addition, the different intracellular compartments also have different biochemical and chemical compositions. The one common characteristic of the different environments is the presence of water. Water, the solvent in which the substances required for the cell's existence are dissolved or suspended, has an important role in the well-being of all cells. The unique physicochemical properties of water make life as we know it possible.

Structure of Water

Two hydrogen atoms share their electrons with an unshared pair of electrons of an oxygen atom to form the water molecule. This deceptively simple molecule, however, has a number of unusual properties. The oxygen nucleus has a stronger attraction for the shared electrons than the hydrogen, and the positively charged hydrogen nuclei are left with an unequal share of electrons, creating a partial positive charge on each hydrogen and a partial negative charge on the oxygen. The bond angle between the hydrogens and the oxygen is 104.5°, making the molecule electrically asymmetric and producing a dipole (Figure 1.2). Water molecules interact because the positively charged hydrogens are attracted to the negatively charged oxygen atom, with the formation of a weak bond between the two water molecules, as in Figure 1.3. This bond is termed a hydrogen bond. Five molecules of water form a tetrahedral structure (Figure 1.3), since each oxygen can share its electrons with four hydrogens and each hydrogen with another oxygen. In solid water, ice, a tetrahedral lattice structure is formed. It is the hydrogen bonding between mol-

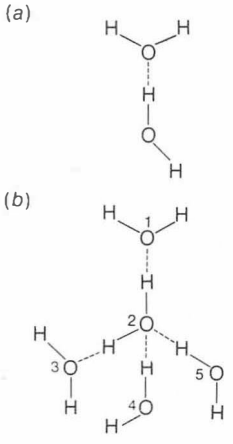


Figure 1.3
(a) Hydrogen bonding between two water molecules.
(b) Tetrahedral hydrogen bonding of five water molecules. Water molecules 1, 2, and 3 are in the plane of the page, 4 is below, and 5 is above.

ecules that gives ice its crystalline structure. Some of these bonds are broken as ice is transformed to liquid water. Each hydrogen bond is relatively weak compared to a covalent bond, but the large number between molecules in liquid water is the reason for the stability of water. Even at 100°C liquid water contains a significant number of hydrogen bonds, which accounts for the high heat of vaporization; in the transformation from a liquid to a vapor state the hydrogen bonds are disrupted.

Water molecules also hydrogen-bond to different chemical structures. Hydrogen bonding also occurs between other molecules and even within a molecule wherever an electronegative oxygen or nitrogen comes in close proximity to a hydrogen covalently bonded to another electronegative group. Some representative hydrogen bonds are presented in Figure 1.4. Intramolecular hydrogen bonding occurs extensively in large macromolecules such as proteins and nucleic acids and is the basis for their structural stability.

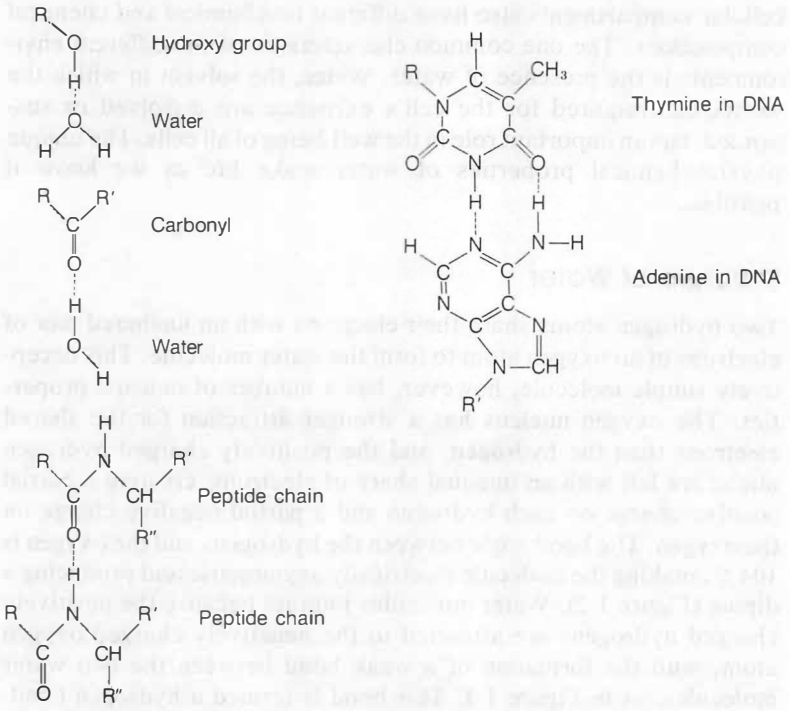


Figure 1.4
Representative hydrogen bonds of importance in biological systems.

Breaking and forming hydrogen bonds occurs more rapidly than covalent bonds because of their low energy; it is estimated that hydrogen bonds in water have a half-life of less than 10^{-10} s. Liquid water actually has a definite structure due to the hydrogen bonding between molecules, but the structure is in a dynamic state as the hydrogen bonds are broken and reformed. A similar dynamic interaction also occurs with substances present in the liquid, which are capable of hydrogen bonding. Many models have been proposed for the structure of liquid water, but none adequately explain all of its properties.

Water as a Solvent

The polar nature of the water molecule and the ability to form hydrogen bonds are the basis for its unique solvent properties. Polar molecules are readily dispersed in water. Salts in which the crystal lattice is held together by the attraction of the positive and negative groups dissolve in water because the electrostatic forces in the crystal can be overcome by the attraction of the charges to the dipole of water. NaCl is an example wherein the electrostatic attraction of the Na^+ and Cl^- is overcome by interaction of Na^+ with the negative charge on the oxygen, and the Cl^- with the positive charge on the protons. A shell of water surrounds the individual ions. The number of weak charge–charge interactions between water and the ions is sufficient to separate the two ions.

Many organic molecules containing nonionic but weakly polar groups are also soluble in water because of the attraction of the groups to water molecules. Sugars and alcohols are readily soluble in water for this reason. Amphipathic molecules, that is, compounds containing both polar and nonpolar groups, will also disperse in water if the attraction of the polar group for water can overcome possible hydrophobic interactions of the nonpolar portions of the molecules. Very hydrophobic molecules, such as those compounds containing long hydrocarbon chains, however, will not disperse in water but do interact with one another to exclude the polar water molecules.

Electrolytes

Substances that dissociate in water into a cation (positively charged ion) and an anion (negatively charged ion) are classified as electrolytes. The presence of these charged ions, prevented from interacting with one another because of the attraction of water molecules to the individual ions, facilitates the conductance of an electrical cur-

rent through an aqueous solution. Sugars or alcohols, which readily dissolve in water but do not carry a charge or dissociate into species with a charge, are classified as nonelectrolytes.

Salts of the alkali metals (e.g., Li, Na, K), when dissolved in water at low concentrations, dissociate totally; at high concentrations there is an increased potential for interaction of the two ions. For biological systems, it is customary to consider that such compounds are totally dissociated. The salts of the organic acids, for example, sodium acetate, also dissociate, but the acetate ion to a limited degree reacts with a proton to form the undissociated acid; these salts are also electrolytes. It is important to remember that when such salts are dissolved in water the individual ions are present in solution rather than the undissociated salt. If a solution has been prepared with several different salts (e.g., NaCl K_2SO_4, and Na acetate) the original molecules do not exist as such in the solution, only the ions (e.g., Na^+, K^+, SO_4^{2-}, and acetate$^-$).

Many compounds, however, when dissolved in water do not totally dissociate but rather establish an equilibrium between the undissociated compound and two or more ions. An example is lactic acid, an important metabolic intermediate, which partially dissociates into a lactate anion and a H^+ as follows:

$$CH_3—CHOH—COOH \rightleftharpoons CH_3—CHOH—COO^- + H^+$$

Because of the partial dissociation, however, such compounds have a lower capacity to carry an electrical charge on a molar basis when compared to a compound that totally dissociates; they are termed weak electrolytes.

Dissociation of Weak Electrolytes

In the partial dissociation of a compound, such as lactic acid, represented by HA, the concentration of the various species where A^- represents the dissociated anion can be determined from the equilibrium equation

$$K'_{eq} = \frac{[H^+][A^-]}{[HA]}$$

The brackets indicate the concentration of each component in moles per liter.

The activities of each species rather than concentration should be employed in the equilibrium equation, but most of the compounds of concern to biochemists are present in low concentrations, and the

value of the activity approaches that for the concentration; the equilibrium constant is indicated as K'_{eq} to indicate that it is an apparent equilibrium constant. The K'_{eq} is a function of the temperature of the system, increasing with increasing temperatures. The degree of dissociation of an electrolyte will depend on the affinity of the anion for a proton; if the weak dipole forces of water interacting with the anion and cation are stronger than the electrostatic forces between the H^+ and anion, there will be a greater degree of dissociation. From the dissociation equation above it is apparent that if the degree of dissociation of a substance is small, K'_{eq} will be a small number, but if the degree of dissociation is large, the number will be large. Obviously, for compounds that dissociate totally, a K'_{eq} cannot be determined because at equilibrium there is no remaining undissociated solute.

Dissociation of Water

Water also dissociates as follows:

$$HOH \rightleftharpoons H^+ + OH^-$$

For convenience, the proton will not be presented in the hydrated form, H_3O^+, even though this is the chemical species actually present. At 25°C the value of K'_{eq} is very small and is in the range of about 1.8×10^{-16}.

$$K'_{eq} = 1.8 \times 10^{-16} = \frac{[H^+][OH^-]}{[H_2O]}$$

With such a small K'_{eq} there is nearly an insignificant dissociation of water, and the concentration of water will be essentially unchanged; the value is 55.5 M. The equation can be rewritten as follows:

$$K'_{eq} \cdot [H_2O] = [H^+][OH^-]$$

The value of $K'_{eq} \times [55.5]$ equals the product of H^+ and OH^- concentrations and is termed the ion product of water. The value at 25°C is 1×10^{-14}. In pure water the concentration of H^+ equals OH^-, and substituting $[H^+]$ for $[OH^-]$ in the equation above, the $[H^+]$ is 1×10^{-7} M. Similarly, the $[OH^-]$ is also 1×10^{-7} M. If either the $[H^+]$ or $[OH^-]$ is varied, a concomitant change in the other ion must occur. Using the equation for the ion product, the $[H^+]$ or $[OH^-]$ can be calculated if the concentration of one of the ions is known.

The importance of the hydrogen ion in biological systems will become apparent in subsequent chapters. For convenience $[H^+]$ is

Table 1.1 Relationship Between $[H^+]$ and pH and $[OH^-]$ and pOH

$[H^+]$ (M)	*pH*	$[OH^-]$ (M)	*pOH*
1.0	0	1×10^{-14}	14
0.1 (1×10^{-1})	1	1×10^{-13}	13
1×10^{-2}	2	1×10^{-12}	12
1×10^{-3}	3	1×10^{-11}	11
1×10^{-4}	4	1×10^{-10}	10
1×10^{-5}	5	1×10^{-9}	9
1×10^{-6}	6	1×10^{-8}	8
1×10^{-7}	7	1×10^{-7}	7
1×10^{-8}	8	1×10^{-6}	6
1×10^{-9}	9	1×10^{-5}	5
1×10^{-10}	10	1×10^{-4}	4
1×10^{-11}	11	1×10^{-3}	3
1×10^{-12}	12	1×10^{-2}	2
1×10^{-13}	13	0.1 (1×10^{-1})	1
1×10^{-14}	14	1.0	0

Table 1.2 pH of Some Biological Fluids

Fluid	*pH*
Blood plasma	7.4
Interstitial fluid	7.4
Intracellular fluid	
Cytosol (liver)	6.9
Lysosmal matrix	5.5–6.5
Gastric juice	1.5–3.0
Pancreatic juice	7.8–8.0
Human milk	7.4
Saliva	6.4–7.0
Urine	5.0–8.0

usually expressed in terms of pH, calculated as follows:

$$\text{pH} = \log \frac{1}{[H^+]}$$

In pure water the hydrogen ion and hydroxy ion are both 1×10^{-7} M, and the pH = 7.0. The OH^- ion concentration can be expressed in a similar fashion as the pOH. For the equation describing the dissociation of water, $1 \times 10^{-14} = [H^+][OH^-]$, and taking the negative logarithm of both sides, the equation becomes 14 = pH + pOH. Table 1.1 presents the relationship between pH and H^+ concentration.

The pH of different biological fluids is presented in Table 1.2. At pH 7, the approximate pH of body fluids, H^+ ion is 0.000,000,1 M (1×10^{-7} M), whereas the concentrations of other cations are between 0.001 and 0.10 M. An increase in H^+ ion concentration to only 0.000,001 (1×10^{-6}) has a marked effect on cellular activities and is deleterious to the continued existence of life; a detailed discussion of the mechanisms by which the body maintains the intra- and extracellular pH's is presented in Chapter 23.

Acids and Bases

The definitions of an acid and a base proposed by Lowry and Brønsted are the most convenient in considering biological systems. (See Figure 1.5.) An *acid* is a *proton donor* and a *base* is a *proton acceptor*. HCl and H_2SO_4 are strong acids because they totally dissociate and OH^- ion is a base because it will accept a proton shifting the equilibrium,

$$OH^- + H^+ \rightleftharpoons H_2O$$

When a strong acid and OH^- are combined, the H^+ from the acid and OH^- interact, participating in an equilibrium with H_2O; since the ion product for water is so small, a neutralization of the H^+ and OH^- occurs.

In dilute solutions strong acids dissociate totally; the anions formed, as an example, Cl^- from HCl, are not classed as bases because they do not accept protons in solution. When an organic acid, such as lactic acid, is dissolved in water, however, it dissociates only partially, establishing an equilibrium between the acid, an anion, and a proton as follows:

$$\text{Lactic acid} \rightleftharpoons \text{lactate}^- + H^+$$

The acid is considered a weak acid and the anion is a base because it

Table 1.3 Some Conjugate Acid–Base Pairs of Importance in Biological Systems

Proton Donor		*Proton Acceptor*
$CH_3—CHOH—COOH$ (lactic acid)	$\rightleftharpoons$	$H^+ + CH_3—CHOH—COO^-$ (lactate)
$CH_3—CO—COOH$ (pyruvic acid)	$\rightleftharpoons$	$H^+ + CH_3—CO—COO^-$ (pyruvate)
$HOOC—CH_2—CH_2—COOH$ (succinic acid)	$\rightleftharpoons$	$2H^+ + {}^-OOC—CH_2—CH_2—COO^-$ (succinate)
$^+NH_3CH_2—COOH$ (glycine)	$\rightleftharpoons$	$H^+ + {}^+NH_3—CH_2—COO^-$ (glycinate)
H_3PO_4	$\rightleftharpoons$	$H^+ + H_2PO_4^-$
$H_2PO_4^-$	$\rightleftharpoons$	$H^+ + HPO_4^{2-}$
HPO_4^{2-}	$\rightleftharpoons$	$H^+ + PO_4^{3-}$
Glucose 6-PO_3H^-	$\rightleftharpoons$	H^+ + glucose 6-PO_3^{2-}
H_2CO_3	$\rightleftharpoons$	$H^+ + HCO_3^-$
NH_4^+	$\rightleftharpoons$	$H^+ + NH_3$
H_2O	$\rightleftharpoons$	$H^+ + OH^-$

will accept a proton reforming the acid. The combination of a weak acid and the base that is formed on dissociation is referred to as a conjugate pair; several examples are presented in Table 1.3. Ammonium ion (NH_4^+) is an acid because it is capable of dissociating to yield a H^+ and ammonia, NH_3, an uncharged species, which is the conjugate base. H_3PO_4 is an acid and PO_4^{3-} is a base, but $H_2PO_4^-$ and HPO_4^{2-} can be classified as either base or acid, depending on whether the phosphate group is accepting or donating a proton.

The tendency of a conjugate acid to dissociate can be evaluated from the K'_{eq}; as indicated above, the smaller the value, the less the tendency to give up a proton and the weaker the acid, the larger a K'_{eq}, the greater the tendency to dissociate a proton, and the stronger the acid.

A convenient method of stating the K'_{eq} is in the form of pK', which is defined as

$$pK' = \log \frac{1}{K'_{eq}}$$

As with pH and $[H^+]$, the relationship between pK' and K'_{eq} is an inverse one, and the smaller the K'_{eq}, the larger the pK'. K'_{eq} and pK's for representative conjugate acids of importance in biological systems are presented in Table 1.4.

Acid = proton donor
Base = proton acceptor

Figure 1.5
Brønsted definition of an acid and base.

Table 1.4 Apparent Dissociation Constant and pK' of Some Compounds of Importance in Biochemistry

Compound		$K'_{eq}M$	pK'
Acetic acid	(CH_3—COOH)	1.74×10^{-5}	4.76
Alanine	(CH_3--CH—COOH)	4.57×10^{-3}	2.34 (COOH)
	NH_3^+	2.04×10^{-10}	9.69 (NH_3^+)
Citric acid	(HOOC—CH_2—COH—CH_2—COOH)	8.12×10^{-4}	3.09
	COOH	1.77×10^{-5}	4.75
		3.89×10^{-6}	5.41
Glutamic	(HOOC—CH_2—CH_2--CH—COOH)	6.45×10^{-3}	2.19 (COOH)
	NH_3^+	5.62×10^{-5}	4.25 (COOH)
		2.14×10^{-10}	9.67 (NH_3^+)
Glycine	(CH_2—COOH)	4.57×10^{-3}	2.34 (COOH)
	NH_3^+	2.51×10^{-10}	9.60 (NH_3^+)
Lactic acid	(CH_3--CHOH—COOH)	1.38×10^{-4}	3.86
Pyruvic acid	(CH_3—CO—COOH)	3.16×10^{-3}	2.50
Succinic	(HOOC—CH_2—CH_2--COOH)	6.46×10^{-5}	4.19
		3.31×10^{-6}	5.48
Glucose 6-PO_3H^-		7.76×10^{-7}	6.11
H_3PO_4		1×10^{-2}	2.0
$H_2PO_4^-$		2.0×10^{-7}	6.7
HPO_4^{2-}		3.4×10^{-13}	12.5
H_2CO_3		1.70×10^{-4}	3.77
NH_4^+		5.62×10^{-10}	9.25

A special case of a weak acid of importance to mammalian cells is that of carbonic acid. Carbon dioxide, when dissolved in water, is involved in the following equilibrium reactions:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

H_2CO_3 is a relatively strong acid with pK<4, but it is in constant equilibrium with physically dissolved CO_2. In an aqueous system in contact with an air phase, dissolved CO_2 is also in equilibrium with CO_2 in the air phase. A change in any of the components in the aqueous phase will cause a shift in both equilibria, for example, increasing CO_2 causes an increase in H_2CO_3, which shifts the equilibrium of the dissociation reaction increasing the H^+. Thus CO_2 is considered a part of the conjugate acid and is in the acid component

of the equation as follows:

$$K'_{eq} = \frac{[H^+][HCO_3^-]}{[H_2CO_3 + CO_2]}$$

Including the CO_2 lowers the K'_{eq} and the value of pK' is 6.1. The actual amount of undissociated H_2CO_3 is less than 1/700 of the CO_2 content and is normally neglected. It is common practice to refer to the dissolved CO_2 as the conjugate acid.

Water is considered a very weak acid, with pK' of 14 at 25°C.

Henderson–Hasselbalch Equation

Changing the concentration of any one component in the equilibrium reaction necessitates a concomitant change in every component. An increase in $[H^+]$ will decrease the concentration of conjugate base with an equivalent increase in the conjugate acid. This relationship is conveniently expressed by rearranging the equilibrium equation and solving for H^+, as shown for the following dissociation:

$$\text{Conjugate acid} \rightleftharpoons \text{conjugate base} + H^+$$

$$K'_{eq} = \frac{[H^+]\,[\text{conjugate base}]}{[\text{conjugate acid}]}$$

Rearranging the equation leads to

$$\frac{1}{[H^+]} = \frac{1}{K'_{eq}} \cdot \frac{[\text{conjugate base}]}{[\text{conjugate acid}]}$$

and, taking the logarithm of both sides,

$$\log \frac{1}{[H^+]} = \log \frac{1}{K'_{eq}} + \log \frac{[\text{conjugate base}]}{[\text{conjugate acid}]}$$

or

$$\text{pH} = \text{p}K' + \log \frac{[\text{conjugate base}]}{[\text{conjugate acid}]}$$

This equation, developed by Henderson and Hasselbalch, is a convenient way of viewing the relationship between the pH of a solution and the relative amounts of base and acid present. Figure 1.6 is a plot of the ratios of conjugate base to conjugate acid on a logarithm scale against the pH for several weak acids. In all cases when the

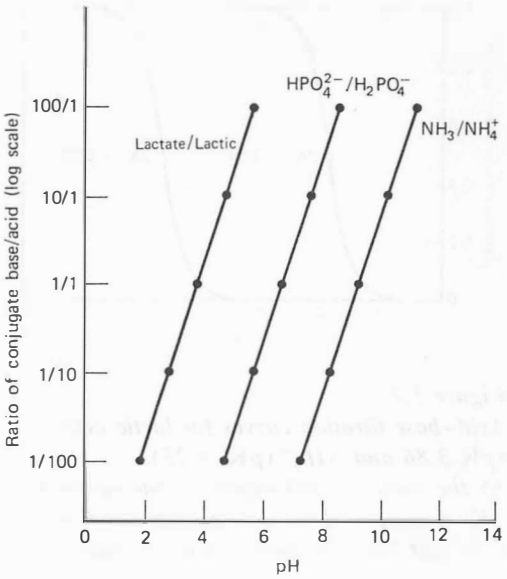


Figure 1.6
Ratio of conjugate base/acid as a function of the pH.
*When the ratio of base/acid is 1, the pH equals the p*K′ *of the weak acid.*

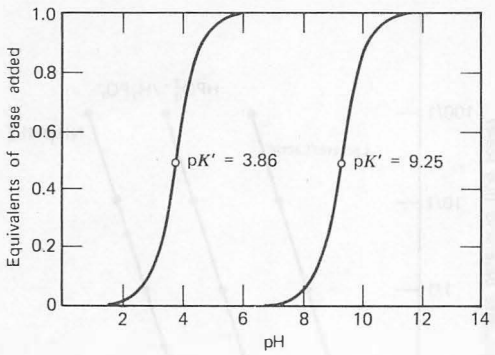


Figure 1.7
Acid–base titration curves for lactic acid (p*K* 3.86 and NH_4^+ (p*K*′ 9.25).
*At the value of pH equal to the respective p*K*′s, there will be an equal amount of the acid and base for each conjugate pair.*

ratio is 1 : 1, the pH equals the pK' of the acid because log 1 = 0 and pH = pK'. If the pH is one unit less than the pK', the base/acid ratio = 1/10, and if the pH is one unit above the pK', the base/acid ratio = 10/1.

Buffers and Buffering

By titrating a weak acid with OH^-, the base/acid ratio can be varied because the OH^- will be neutralized by H^+, decreasing the $[H^+]$ and causing a dissociation of an amount of the weak acid equal to the OH^- added. This will produce an equal increase in the concentration of the conjugate base. Titration curves of several weak acids are presented in Figure 1.7. When 0.5 equiv OH^- are added, 50% of the weak acid is dissociated and the acid/base ratio is 1.0, and the pH at this point is equal to the pK' of the acid. The shapes of the individual titration curves are similar but displaced due to the differences in pK's. As OH^- ion is added, initially there is a rather steep rise in the pH, but between 0.1 and 0.9 equiv OH^-, the pH change is only ~2. Thus a large amount of OH^- is added with a relatively small change in pH.

The best buffering range for a conjugate pair is in the pH range near the pK' of the weak acid. Starting from a pH one unit below to a pH one unit above the pK', ~82% of a weak acid in solution will dissociate, and therefore an amount of base equivalent to about 82% of the original acid can be neutralized with a change in pH of 2. The maximum buffering range for a conjugate pair is considered to be between 1 pH unit above and below the pK. A weak acid such as lactic acid with pK' = 3.86 is an effective buffer in the range of pH 3–5 but has little buffering ability at intracellular pH's between 6.8 and 7.4. The $HPO_4^{2-}/H_2PO_4^-$ pair with pK' = 6.7, however, is an effective intracellular buffer.

The buffering capacity also depends on the concentration of the acid and base pair. A case in point is blood plasma at pH 7.4. The pK' for $HPO_4^{2-}/H_2PO_4^-$ of 6.7 would suggest that this conjugate pair would be an effective buffer; the concentration of the phosphate pair, however, is low compared to the HCO_3^-/CO_2 system with a pK' of 6.1, which is present at a 20-fold higher concentration and accounts for most of the buffering capacity. In considering the buffering capacity both the pK' and the concentration of the conjugate pair must be taken into account. Most organic acids are relatively unimportant as buffers in cellular fluids because their pK's are more than several pH units from the pH of the cell, and their concentrations are too low in comparison to such buffers as the various phosphate pairs and the HCO_3^-/CO_2 system.

1.3 ORGANIZATION AND COMPOSITION OF EUCARYOTIC CELLS

In many cases the membranes of eucaryotic cells observed by electron microscopy (Figure 1.8) define specific subcellular organelles, such as the nucleus, mitochondria, or lysosomes. These are self-contained units and can be isolated essentially intact from cells and tissues. Other cellular membranes are part of a tubular-like network throughout the cell, enclosing an interconnecting space or cisternae. Such is the case for the endoplasmic reticulum and the Golgi complex. On disruption of the cell, these membrane systems are dis-

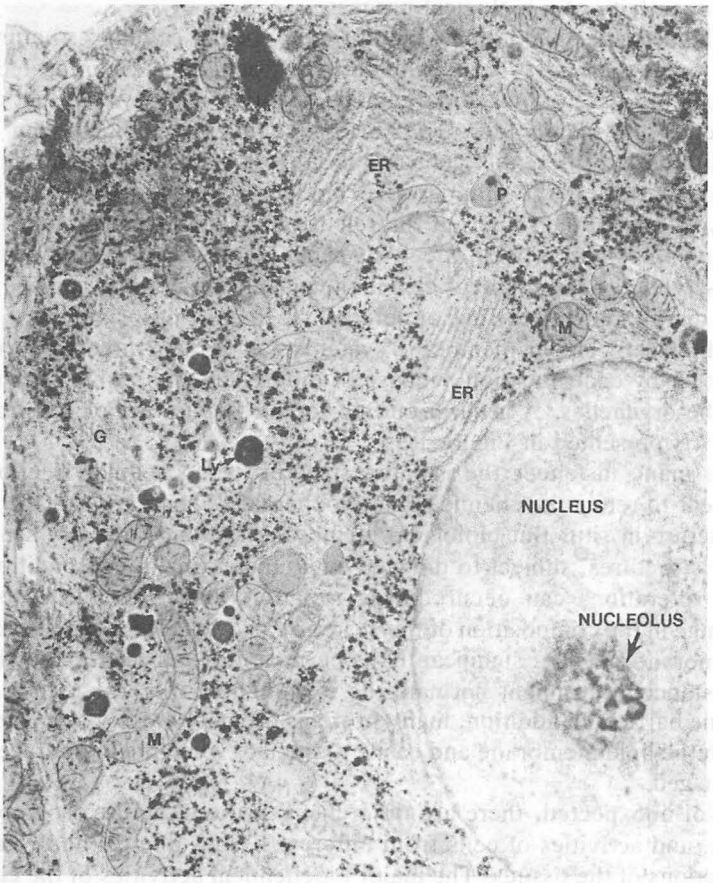


Figure 1.8
An electron micrograph of a rat liver cell labeled to indicate the major structural components of eucaryotic cells.
Note the number and variety of subcellular organelles and the network of interconnecting membranes enclosing channels, that is, cisternae. All eucaryotic cells are not as complex in their appearance, but most contain the major structures shown in this figure.
Photograph reprinted with permission of Dr. K. R. Porter.

rupted, and the membranes reform into small vesicles encapsulating a portion of the isolation medium.

The semipermeable nature of cellular membranes prevents the ready diffusion of many molecules, particularly electrolytes, from one side to another. Very specific transport mechanisms in some membranes for the translocation of both charged and uncharged molecules allow the membrane to modulate the concentrations of the transported substances in various cellular compartments. In addition, macromolecules such as enzymes and nucleic acids do not move readily through biological membranes because of the controls exerted by cellular membranes. The fluid matrix of the various cellular compartments have a distinctive chemical composition. Partitioning of activities and components in the membrane compartments and organelles has a number of advantages to the economy of the cell, including the sequestering of substrates and cofactors, where they are required, and adjustments of pH and ionic composition for maximum activity of a biological process.

The activities and composition of many of the cellular structures and organelles have been defined by histochemical staining techniques of intact tissues and cells and isolation of the individual organelles and membranes. The various cellular components can be separated by differential centrifugation, following disruption of the plasma membrane. A number of techniques are available to alter the plasma membrane to permit the release of subcellular components, including osmotic shock of cell suspensions, homogenization of tissues, where shearing forces cleave the plasma membrane, and chemical disruption with the use of detergents. In an appropriate isolation medium the cellular membrane systems can be separated from one another by centrifugation because of differences in size and density of the organelles. A general outline for the separation of cell fractions is presented in Figure 1.9.

In many instances the isolated structures and cellular fractions appear to retain the chemical and biochemical characteristics of the structure in situ. But biological membrane systems are very sensitive structures, subject to damage even under very mild conditions and alterations can occur during isolation, which can lead to a change in the composition of the structure. The slightest damage to a membrane can alter significantly its permeability properties allowing substances that might normally be excluded to traverse the membrane barrier. In addition, many proteins are only loosely associated with the lipid membrane and easily dissociate when the membrane is damaged.

Not unexpected, there are differences in the structure, composition, and activities of cells from different tissues due to the diverse functions of the tissues. The major biochemical activities of the cell-

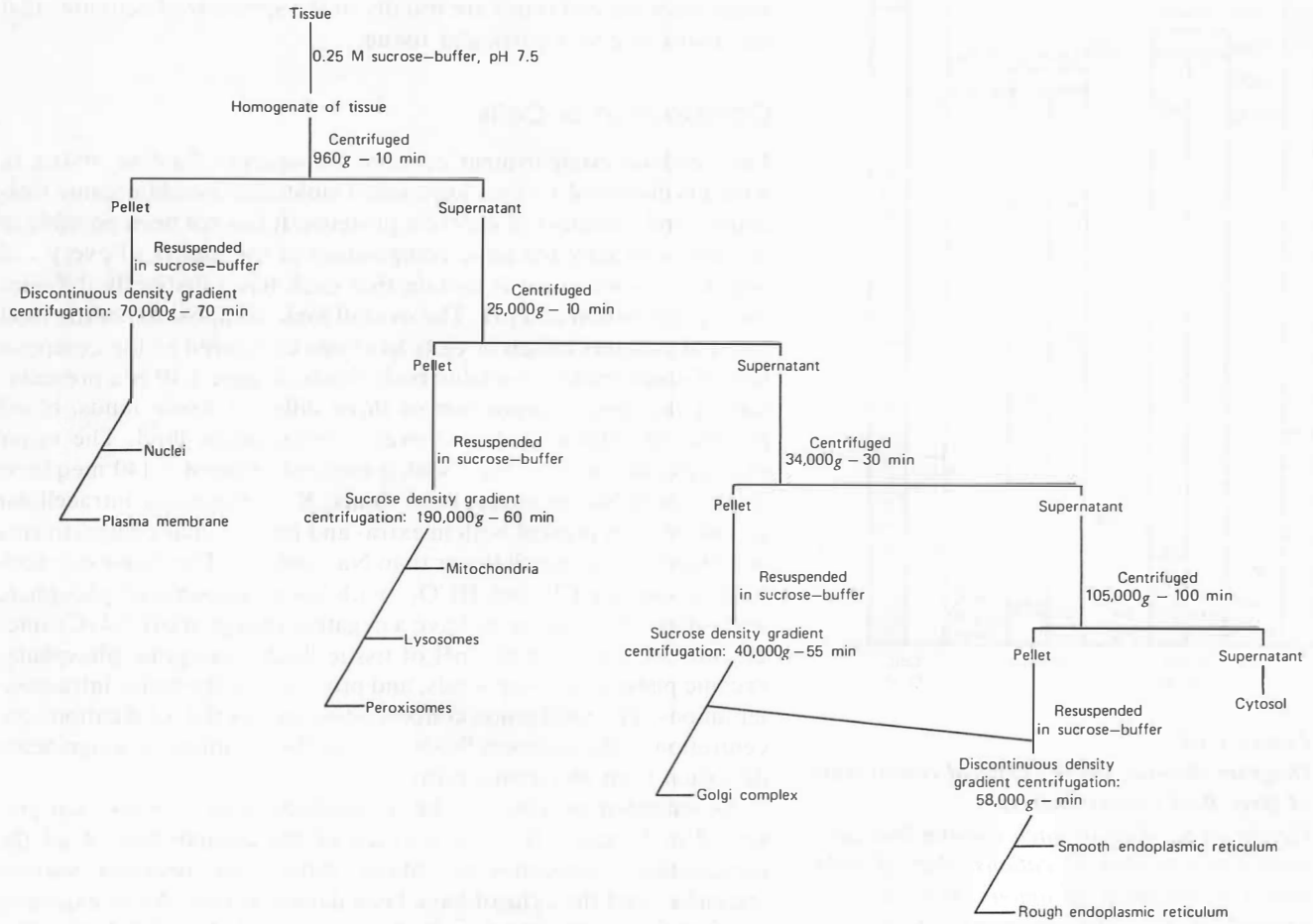


Figure 1.9

Diagram of a general procedure for the isolation of subcellular components from rat liver by differential and density gradient centrifugation. *This is an oversimplified scheme; to simplify the diagram, the composition of the various buffers and of the density gradients are not included. The gravitational fields and times are indicated only to indicate the range of each required for the isolation. There are steps in the procedure, particularly the washing of pellets, that have also been omitted.*

Details of the procedure are found in the article by S. Fleischer and M. Kervina, Subcellular fractionation of rat liver. *Methods Enzymol.*, 31:6, 1974.

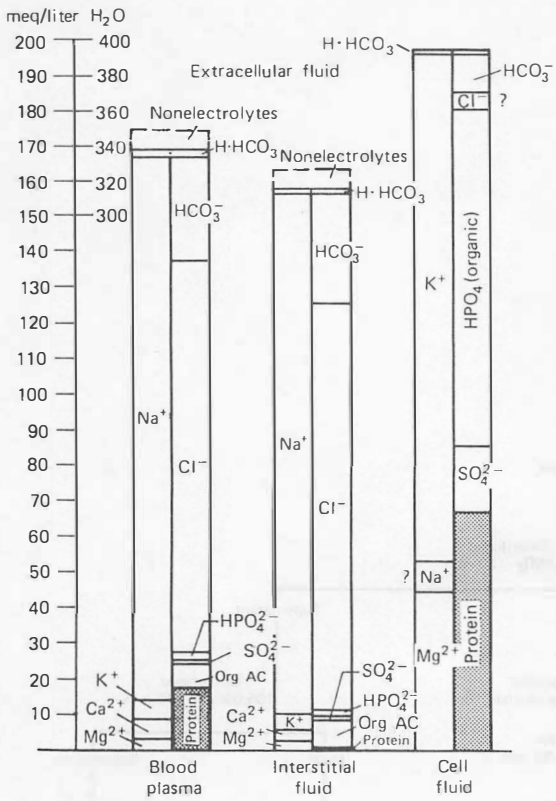


Figure 1.10

Diagram showing major chemical constituents of three fluid compartments.

Height of left half of each column indicates total concentration of cations; that of right half, concentration of anions. Both are expressed in meq/liter of water. Note that chloride and sodium values in cell fluid are questioned. It is probable that, at least in muscle, the cytosol contains some sodium but no chloride.

Modified from Gamble. From Magnus I. Gregersen, in *Medical Physiology*, 11th ed., Philip Bard, ed., Mosby, St. Louis, MO. 1961, p. 307.

ular organelles and membrane systems, however, are fairly constant from tissue to tissue. Thus, results of a study of biochemical pathways in liver are most often applicable to other tissues. The differences between cell types are usually in the specialized activities that are distinctive to a particular tissue.

Composition of Cells

Each cellular compartment contains an aqueous fluid or matrix in which is dissolved various ions, small molecular weight organic molecules, and a number of different proteins. It has not been possible to determine exactly the ionic composition of the matrix of every cell organelle; however, it is certain that each has a distinctly different ionic composition and pH. The overall ionic composition of the total internal aqueous milieu of cells has been compared to the composition of other readily available body fluids. Figure 1.10 is a presentation of the ionic composition of three different tissue fluids, blood plasma, interstitial fluid, and overall intracellular fluid. The major extracellular cation is Na^+, with a concentration of ~140 meq/liter; there is little Na^+ in intracellular fluids. K^+ is the major intracellular cation. Mg^{2+} is present both in extra- and intracellular compartments at concentrations much lower than Na^+ and K^+. The major extracellular anions are Cl^- and HCO_3^- with lower amounts of phosphate and sulfate. Most proteins have a negative charge at pH 7.4 (Chapter 2), thus are anions at the pH of tissue fluids. Inorganic phosphate, organic phosphate compounds, and proteins are the major intracellular anions. The total anion concentration equals the total cation concentration in the different fluids, in that there cannot be a significant deviation from electroneutrality.

As indicated previously, the intracellular ionic composition presented in Figure 1.10 is an average of the composition of all the intracellular compartments. Major differences between various organelles and the cytosol have been demonstrated. As an example, the free Ca^{2+} concentration in the cytosol is below 0.001 meq/liter but is much higher in the cisternae of the sarcoplasmic reticulum and in mitochondria. These two membrane systems in muscle actually control the cytosolic concentration of Ca^{2+} by actively sequestering the cellular Ca^{2+}; release of Ca^{2+} into the cytosol triggers a number of biochemical reactions and muscle contractions.

The concentration of most small molecular weight organic molecules, such as sugars, organic acids, amino acids, and phosphorylated intermediates, all of which serve as substrates for enzymatic reactions in the various cellular compartments, is considered to be in the range of 0.01–1.0 mM, but in some cases they have significantly lower concentrations, depending on the cell and individual organelle.

Coenzymes, organic molecules required for the maximum activity of different enzymes, are in the same range of concentration. In contrast to inorganic ions, substrates are present in relatively low concentrations when considering their overall cellular concentrations but can be increased by localization in a specific organelle or cellular microenvironment.

It is not very meaningful to determine the molar concentration of individual proteins in cells. In many cases they are localized with specific structures or in combination with other proteins to create a functional unit. It is in a restricted compartment of the cells that the individual proteins carry out their role, whether structural, catalytic, or regulatory.

1.4 FUNCTIONAL ROLE OF SUBCELLULAR ORGANELLES AND MEMBRANES

The subcellular localization of various metabolic pathways will be described throughout this textbook. In some cases an entire pathway is located in a single cellular compartment, but many metabolic sequences are divided between two locations, with the intermediates in the pathway moving or being translocated from one cell compartment to another. In general, the organelles have very specific functions. The specific activity of particular organelles, such as mitochondria or lysosomes, can be employed as an identifying characteristic during isolation.

The following describes briefly some of the major roles of the eucaryotic cell structures indicated in Figure 1.8; the descriptions are not inclusive but rather are meant to indicate the level of complexity and organization of the cell. Details of some cellular components are described in later chapters during discussions of their function in metabolism.

Plasma Membrane

The limiting membrane of every cell has a unique role in the maintenance of a cell's integrity. One surface is in contact with the variable external environment, and thus it is responsible for excluding many substances but permitting the entrance of others. A major activity of the plasma membrane is the mediated transport of ions and compounds both in and out of the cell. It is also responsible for cell

recognition and cellular movements. Plasma membranes from a variety of cells have been isolated and extensively studied; details of their structure and biochemistry are presented in Chapter 5.

Nucleus

The early microscopists divided the cell into a nucleus, the largest membrane bound compartment, and the cytoplasm, the remainder of the cell. The nucleus is surrounded by two membranes, termed the perinuclear envelope, with the outer membrane being continuous with membranes of the endoplasmic reticulum. The nucleus contains a subcompartment, clearly seen in electron micrographs, termed the nucleolus. The vast majority of cellular deoxyribonucleic acid (DNA) is located in the nucleus in the form of a DNA–protein complex termed chromatin. Chromatin is organized into chromosomes. DNA is the repository of the cell's genetic information. The importance of the nucleus in cell division and for controlling the phenotypic expression of genetic information is well established. The biochemical reactions involved in the replication of DNA during mitosis and the repair of DNA following damage (Chapter 17), the transcription of the information stored in DNA into a form that can be translated into proteins of the cell (Chapter 18), are contained in the nucleus. Transcription of DNA involves the synthesis of ribonucleic acid (RNA), which is processed following synthesis into a variety of forms. Part of this processing occurs in the nucleolus, which is very rich in RNA. The nucleus may synthesize some of the proteins required for nuclear function, but this activity is insignificant in comparison to the protein synthetic activity of the endoplasmic reticulum.

Endoplasmic Reticulum

The cytoplasm of most eucaryotic cells contains a network of interconnecting membranes enclosing channels, that is, cisternae, that thread from the perinuclear envelope of the nucleus to the plasma membrane. This extensive subcellular structure, termed the endoplasmic reticulum, consists of membrane structures with a rough appearance in some areas and smooth in other places. The rough appearance is due to the presence of ribonucleoprotein particles, that is, ribosomes, attached on the cytoplasmic side of the membrane. Smooth endoplasmic reticulum does not contain ribosomal particles. During cell fractionation the endoplasmic reticulum network is disrupted, with the membrane resealing into small vesicles referred to as microsomes, which can be isolated by differential centrifugation. Microsomes per se do not occur in cells.

The major function of the ribosomes on the rough endoplasmic reticulum is the biosynthesis of proteins for export to the outside of the cell. The endoplasmic reticulum also contains enzymes involved in the biosynthesis of steroid hormones and a variety of oxidative and transferase reactions required for removal of toxic substances. The endoplasmic reticulum also has a role with the Golgi complex in the formation of other cellular organelles such as lysosomes and peroxisomes.

Golgi Complex

Only in the last few decades has the biochemical role of the Golgi complex been clearly defined. This network of flattened smooth membranes and vesicles is responsible for the secretion to the external environment of a variety of proteins synthesized on the endoplasmic reticulum. Golgi membranes catalyze the transfer of glycosyl groups to proteins, a chemical modification required for transport of proteins across the plasma membrane. In addition, the complex is a major site of new membrane formation. Membrane vesicles are formed from the Golgi complex in which various proteins and enzymes are encapsulated, which can be secreted from the cell after an appropriate signal. The digestive enzymes synthesized by the pancreas are stored in intracellular vesicles formed by the Golgi complex and released by the cell when needed in the digestive process. The role in membrane synthesis also includes the formation of intracellular organelles such as lysosomes and peroxisomes.

Mitochondria

Mitochondria have been studied extensively because of their role in cellular energy metabolism and the ease with which they can be isolated from tissues in a relatively intact state. In electron micrographs mitochondria appear as spheres, rods, or filamentous bodies; they are usually about 0.5–1 μm in diameter and up to 7 μm in length. The internal matrix is surrounded by two membranes, distinctively different in appearance and biochemical function. The inner membrane convolutes into the matrix of the mitochondrion to form cristae and contains numerous small spheres attached by stalks on the inner surface. The outer and inner membranes contain distinctly different sets of enzymes, which are used for identification of mitochondria during isolation. The components of the respiratory chain and the mechanism for ATP synthesis are part of the inner membrane, described in detail in Chapter 6. In addition to membrane-bound enzymes, the space between the two membranes and the internal matrix also contains a variety of enzymes. Major

Table 1.5 Representative Lysosomal Enzymes and Their Substrates

Type of Substrate and Enzyme	*Specific Substrate*
Polysaccharide-hydrolyzing enzymes	
α-Glucosidase	Glycogen
α-Fucosidase	Membrane fucose
β-Galactosidase	Galactosides
α-Mannosidase	Mannosides
β-Glucuronidase	Glucuronides
Hyaluronidase	Hyaluronic acid and chondroitin sulfates
Arylsulfatase	Organic sulfates
Lysozyme	Bacterial cell walls
Protein-hydrolyzing enzymes	
Cathepsins	Proteins
Collagenase	Collagen
Elastase	Elastin
Peptidases	Peptides
Nucleic acid-hydrolyzing enzymes	
Ribonuclease	RNA
Deoxyribonuclease	DNA
Lipid-hydrolyzing enzymes	
Esterase	Fatty acid esters
Phospholipase	Phospholipids
Phosphatases	
Phosphatase	Phosphomonoesters
Phosphodiesterase	Phosphodiesters

metabolic pathways involved in the oxidation of carbohydrates, lipids, and amino acids, and special biosynthetic pathways involving urea and heme synthesis are located in the mitochondrial matrix space. The outer membrane is relatively permeable, but the inner membrane is highly selective and contains a number of transmembrane transport systems. The inner membrane contains a specific transporter to move Ca^{2+} into and out of the matrix of the mitochondria and it is proposed that mitochondria have a role in the maintenance of cytoplasmic Ca^{2+} levels.

Mitochondria also contain a specific DNA, containing genetic information for some of the mitochondrial proteins, and the biochemical equipment for limited protein synthesis. The presence of this biosynthetic capacity indicates the unique role that mitochondria have in their own destiny.

Lysosomes

Digestion of a variety of substances inside the cell occurs in the structures designated as lysosomes. These cellular organelles have a single limiting membrane, capable of maintaining a significant pH gradient between the lysosomal matrix and the cytosol. Encapsulated in lysosomes is a class of enzymes termed hydrolases, which catalyze the hydrolytic cleavage of carbon–oxygen, carbon–nitrogen, carbon–sulfur, and oxygen–phosphorus bonds in proteins, lipids, carbohydrates, and nucleic acids. As in the process of digestion in the lumen of the gastrointestinal system, the enzymes of the lysosome are able to split molecules into simple low molecular weight compounds, which can be utilized by the metabolic pathways of the cell. The enzymes of the lysosome have a common characteristic in that each is most active when the pH of the medium is acidic. The relationship between pH and enzyme activity is discussed in Chapter 4. The pH of the cytosol is close to neutral, pH 7.0 and the lysosomal enzymes have little activity at this pH. Thus for the lysosomal enzymes to carry out the digestion of various substances the intralysosomal pH must be significantly lower than the cystosolic pH. A partial list of the enzymes present in lysosomes is presented in Table 1.5.

The individual lysosomal enzyme content of different tissues varies and apparently depends on specific needs of individual tissues to digest different substances. Intact lysosomes isolated from other cellular components do not catalyze the hydrolysis of substrates until the membrane is disrupted, demonstrating that the lysosomal membrane is a barrier to the ready access of cellular components to the interior of the lysosome. The lysosomal membrane can be disrupted by various treatments, leading to a release of the lysosomal en-

zymes, which can then react with their individual substrates. The activities of the lysosomal enzymes are termed "latent" to indicate the need to disrupt the membrane to determine the activity. Disruption of the membrane in situ can lead to cellular digestion; a variety of pathological conditions have been attributed to release of lysosomal enzymes, including arthritis, allergic responses, several muscle diseases, and drug-induced tissue destruction. (See Figure 1.11.)

Lysosomes are involved in the normal digestion of both intra- and extracellular substances that must be removed by the cell. By the process of endocytosis, external material is taken into the cell and encapsulated in a membrane-bound vesicle. Formed foreign substances such as microorganisms are engulfed by the cell membrane by the process of phagocytosis, and extracellular fluid containing suspended material is taken up by pinocytosis. In both processes the vesicle containing the external material interacts with a lysosome to form a cystolic organelle containing both the material to be digested and the enzymes capable of carrying out the digestion. These vacuoles are identified microscopically by their size and often by the presence of partially formed structures in the process of being digested. Lysosomes in which the enzymes are not as yet involved in the digestive process are termed primary lysosomes, whereas secondary lysosomes are organelles in which digestion of material is underway. The latter are also referred to as digestive vacuoles. The general sequence of digestion of extracellular substances is represented in Figure 1.12.

Cellular constituents are turning over continuously, and there is a constant degradation of normal components. Lysosomes also have the responsibility of digesting the cellular debris. The dynamic synthesis and degradation of cellular substances includes proteins and nucleic acids, as well as structures such as mitochondria and the endoplasmic reticulum. During the normal self-digestion process, that is, autolysis, cellular substances are encapsulated within a membrane vesicle that reacts with a lysosome to complete the degradation. The overall process is termed autophagy and is also represented in Figure 1.12.

The products of the normal lysosomal digestive process are apparently able to diffuse across the lysosomal membrane and be reutilized by the cell. Indigestible material, however, accumulates in vesicles referred to as residual bodies; the content of these vesicles are removed from the cell by exocytosis, where the membrane of a vesicle interacts with the plasma membrane. In some cases, however, the residual bodies, which contain a high concentration of lipid, persist for long periods of time. The lipid is oxidized and a pigmented substance, which is chemically heterogeneous, contain-

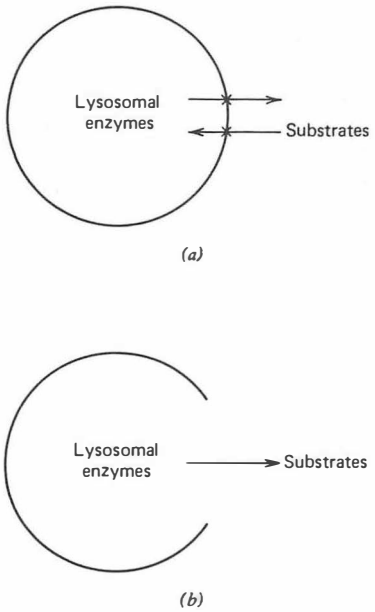


Figure 1.11
Latency of lysosomal enzymes.
(a) Intact lysosomes: enzymes inactive with external substrates.
(b) Disrupted lysosomes: enzymes active with external substrates.
When the membrane of the lysosome is intact, substrates that are external will not react with the intralysosomal enzymes. Disruption of the membrane by physical or chemical means leads to a release of the enzymes and the hydrolysis of external substrates.

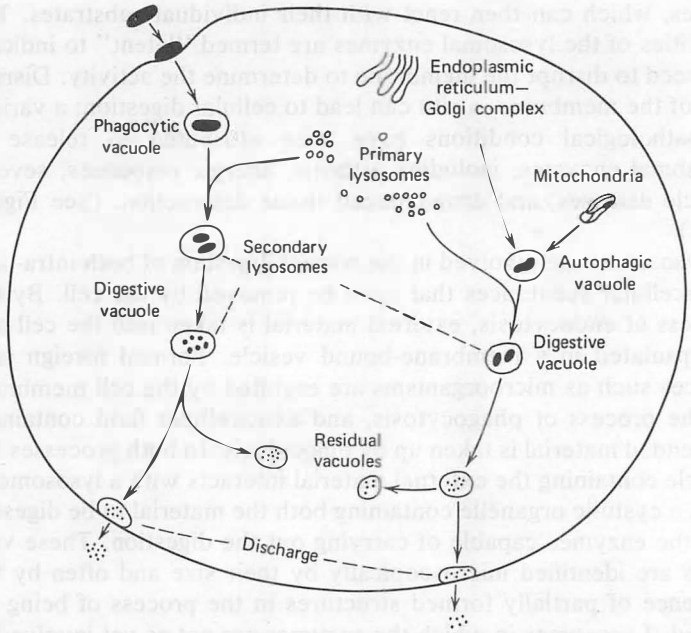


Figure 1.12
Diagrammatic representation of the role of lysosomes in the intracellular digestion of substances internalized by phagocytosis (heterophagy) and of cellular components (autophagy).
In both processes the substances to be digested are enclosed in a membrane vesicle, which is followed by interaction with either a primary or secondary lysosome.

ing polyunsaturated fatty acids and protein, accumulates in the cell. This material, termed lipofuscin, has been called the "age pigment" or the "wear and tear pigment" because it accumulates in cells of older individuals. It occurs in all cells but particularly in neurons and muscle cells and has been implicated as having a role in aging.

Lysosomal enzymes in some cells also participate in the secretory process by hydrolyzing specific bonds in precursor protein molecules, leading to the formation of active proteins that are secreted from the cell. Under controlled conditions lysosomal enzymes are secreted from the cell for the digestion of extracellular material; an extracellular role for some lysosomal enzymes has been demonstrated in connective tissue and prostate gland, and in the process of morphogenesis.

The absence of specific lysosomal enzymes have been demonstrated in a number of genetic diseases; in these cases there is an accumulation in the cell of specific cellular components that cannot be digested. Lysosomes of the effected cell become enlarged with undigested material, which interferes with normal cellular processes. A discussion of lysosomal storage diseases is presented in Chapter 10.

Peroxisomes

Most eucaryotic cells of mammalian origin, as well as protozoan and plants, have a defined cellular organelle which contains several enzymes that either produce or utilize hydrogen peroxide. Frequently referred to as microbodies, the designation peroxisome is now more widely accepted. The organelles are small (0.3–1.5 μm in diameter), spherical or oval in shape, with a granular matrix and in some cases a crystalline inclusion termed a nucleoid. Peroxisomes contain enzymes that oxidize D-amino acids, uric acid, and various 2-hydroxy acids; in each case hydrogen peroxide is formed as a product of the reaction. The enzyme catalase, which catalyzes the conversion of two hydrogen peroxides to water and oxygen, is also present in the organelle. It has been proposed that by having both the peroxide-producing and -utilizing enzymes in one cellular compartment the cell protects itself from the toxicity of hydrogen peroxide.

Microtubules and Microfilaments

Eucaryotic cells contain both microtubules and microfilaments, which participate in cell division, maintenance of morphology, intracellular transport, and cell motility. The microtubules, which consist of a polymer of the protein tubulin, can be rapidly assembled and disassembled, depending on the needs of the cell. A very important cellular filament is in striated muscle and is responsible for muscular contraction (see Chapter 22). Microtubules and filaments do not have a specific role in cellular metabolism but are important in the structural aspects of the cell.

Cytosol

The least complex in structure, but not in chemistry, is the remaining matrix or cytosol of the cell. It is here that many of the multiplicity of chemical reactions of metabolism occur and where substrates and cofactors of various enzymes interact. Even though there is no apparent structure to the cytoplasm, the high protein content precludes the matrix from being a truly homogeneous mixture of soluble

components. Many reactions may be localized in selected areas of the cell, where the conditions of substrate availability are more favorable for the reaction. The actual physicochemical state of the cytosol is poorly understood. A major role of the cytosol is to support the synthesis of proteins catalyzed by the rough endoplasmic reticulum by supplying cofactors and enzymes. In addition, the cytosol contains free ribosomes, often in a polysome form, for synthesis of intracellular proteins.

Studies with isolated cell cytosol suggest that many reactions are catalyzed by soluble enzymes, but in the intact cell some of these enzymes might be loosely attached to one of the many membrane structures and released upon cell disruption.

Conclusion

The eucaryotic cell is a complex machine whose purpose is to replicate itself when necessary, maintain an intracellular environment to permit a myriad of complex reactions to occur as efficiently as possible, and to protect itself from the hazards of its surrounding environment. The cells of multicellular organisms also participate in maintaining the well-being of the whole organism by exerting influences on each other to maintain all tissue and cellular activities in balance. Thus, as we dissect the separate chemical components and activities of cells, it is important to keep in mind the concurrent and surrounding activities, constraints, and influences. Only by bringing together all the separate parts, that is, reassembling the puzzle, will we appreciate the wonder of a living cell.

BIBLIOGRAPHY

Water and Electrolytes

Dick, D. A. T. *Cell water*. Washington, D.C.: Butterworths, 1966.

Eisenberg, D., and Kauzmann, W. *The structures and properties of water*. Fairlawn, N.J.: Oxford University Press, 1969.

Morris, J. G. *A biologist's physical chemistry*. Reading, Mass.: Addison-Wesley, 1968.

Cell Structure

Alterman, P. L., and Katz, D. D. (eds.). *Cell biology*. Bethesda, Md.: Federated American Society of Experimental Biology, 1975.

Fawcett, D. W. *The cell: its organelles and inclusions*. Philadelphia: W. B. Saunders, 1966.

Flickinger, C. J., Brown, J. C., Kutchai, H. C., and Ogilvie, J. W. *Medical cell biology*. Philadelphia: W. B. Saunders, 1979.

Hers, H. G., and Van Hoof, F. (eds.). *Lysosomes and storage diseases*. New York: Academic, 1973.

Loewy, A., and Siekevitz, P. *Cell structure and function*, 2nd ed. New York: Holt, Rinehart & Winston, 1970.

Porter, K. R., and Bonneville, M. A. *Fine structure of cells and tissues*. Philadelphia: Lea & Febiger, 1972.

Proteins I

RICHARD M. SCHULTZ

COMPOSITION AND STRUCTURE

2.1 FUNCTIONAL ROLES OF PROTEINS IN HUMANS

Proteins perform a surprising variety of essential functions in mammalian organisms. These functions may be grouped into two classes: dynamic and structural (static). Dynamic functions of proteins include transport, metabolic control, contraction, and catalysis of chemical transformations. In their structural functions, proteins provide the matrix for bone and connective tissue, giving structure and form to the human organism.

One of the important groups of proteins that have a dynamic type of function are the enzymes. Enzymes act to catalyze chemical reactions. Almost all the thousands of different chemical reactions that occur in living organisms and involve covalent bond formation or cleavage require a specific enzyme catalyst for the reaction to occur at a rate compatible with life. Thus the characteristics and functions of any cell are based on the particular chemistry of the cell, which in turn is generated by the specific enzyme makeup of the cell. Genetic traits are expressed through the synthesis of enzymes, which catalyze the chemical reactions that establish the trait to be expressed.

A second important dynamic type of function for proteins is transport. Particular examples discussed in greater detail in this text are hemoglobin and myoglobin, which transport oxygen in blood and in muscle, respectively. Transferrin transports iron in blood. Other important proteins act to transport hormones in blood from their site of synthesis to their site of action. Many drugs and toxic compounds are transported bound to proteins.

Proteins can also function, dynamically, in a protective role. For example, the immunoglobulins and interferon are proteins that act against bacterial or viral infection. Fibrin is a protein that is formed where required to stop the loss of blood on injury to the vascular system.

Many hormones are proteins. Protein hormones include insulin, thyrotropin, somatotropin (growth hormone), luteinizing hormone, and follicle stimulating hormone. There are many diverse protein-type hormones that have a low molecular weight (mol $<5{,}000$), referred to as *peptides*. In general, the term protein is used for molecules composed of over 50 component amino acids and the term peptide is used for molecules of less than 50 amino acids. Important peptide hormones include adrenocorticotropin, antidiuretic hormone, glucagon, and calcitonin.

Some proteins have roles in the contractile mechanism. Of particular importance are the proteins myosin and actin, which function in muscle contraction.

Proteins are active in the control and regulation of gene transcription and translation. These include the histone proteins associated with DNA, the repressor proteins that control gene expression, and the proteins that form a part of the ribosomes.

Whereas the above proteins are "dynamic" in their function, other proteins have structural, "brick-and-mortar" roles. This group of structurally functional proteins includes collagen and elastin, which form the matrix for bone and ligaments and provide structural strength and elasticity to the organs and the vascular system. α-Keratin has an essential structural role in epidermal tissue.

It is obvious that an understanding of both the normal functioning and the pathology of the mammalian organism requires a clear understanding of the structure and properties of the proteins.

2.2 AMINO ACID COMPOSITION OF PROTEINS

Proteins Are Polymers of α-Amino Acids

It is amazing that all the different types of proteins are initially synthesized as polymers of only 20 amino acids, known as the common amino acids. The *common amino acids are defined as those amino acids for which a specific codon exists in the DNA genetic code.* There are 20 amino acids for which DNA codons are known. The process of the reading of the DNA code, resulting in the polymerization of amino acids of a specific sequence into proteins based on the DNA code, is the basis of a later chapter (Chapter 19). In this chapter we will discuss only the protein product of this genetically controlled synthetic process.

In addition to the common amino acids, derived amino acids are found in proteins. *Derived amino acids are formed from one of the common amino acids,* usually by an enzyme-facilitated reaction, *after the common amino acid has been incorporated into a protein structure.* An example of a derived amino acid is cystine (see the section entitled "Derived Amino Acid Cystine" below). Other derived amino acids are desmosine and isodesmosine found in the protein elastin (Chapter 3), hydroxyproline and hydroxylysine found in collagen (Chapter 3), and γ-carboxyglutamate found in prothrombin.

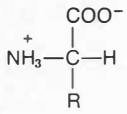


Figure 2.1
General structure of the common amino acids.

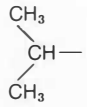
General Structure of the Common Amino Acids

The common amino acids have the general structure depicted in Figure 2.1. They contain in common a central (*alpha*) carbon atom to which a carboxylic acid group, an amino group, and a hydrogen atom are covalently bonded. In addition, the *alpha* (α)-carbon atom binds a side chain group, designated R, that is different for each of the 20 common amino acids.

In the structure depicted in Figure 2.1 the ionized form for a common amino acid that is present in solution at pH 7 is shown. The α-amine is protonated and in its ammonium ion form; the carboxylic acid group is in its unprotonated or carboxylate form.

Side Chain Structures of the Common Amino Acids

The structures for the common amino acids are shown in Figure 2.2. In the category of alkyl amino acids are glycine, alanine, valine, leucine, and isoleucine. Glycine has the simplest structure, with R = H. Alanine contains a methyl (CH_3—) R group and valine an isopropyl R group. The leucine and isoleucine R groups are butyl alkyl chains that are structural isomers of each other. In leucine the branching methyl group in the isobutyl side chain occurs on the *gamma* (γ)-carbon of the amino acid. In isoleucine the butyl side chain is branched at the β-carbon.



Isopropyl R group

In the category of aromatic amino acids are phenylalanine, tyrosine, and tryptophan. In phenylalanine the R group contains a benzene ring, tyrosine contains a phenol group, and the tryptophan R group contains a heterocyclic structure known as an indole. In the three aromatic amino acids the aromatic moiety is attached to the α-carbon through a methylene (—CH_2—) carbon (Figure 2.2).

The two sulfur-containing common amino acids are cysteine and methionine. In cysteine the R group is thiolmethyl ($HSCH_2$—). In methionine the side chain is a methyl ethyl thiol ether ($CH_3SCH_2CH_2$—).

There are two hydroxy (alcohol)-containing common amino acids, serine and threonine. In serine the side chain is a hydroxylmethyl moiety ($HOCH_2$—). In threonine an ethanol structure is connected to the α-carbon of the amino acid through the 1 position of the ethanol, resulting in a secondary alcohol structure for R (CH_3CHOH—).

In proline the side chain group, R, is unique in that it incorporates the α-amino group in the side chain. Thus proline is more accurately classified as an α-imino acid, since its α-amine is a secondary amine,

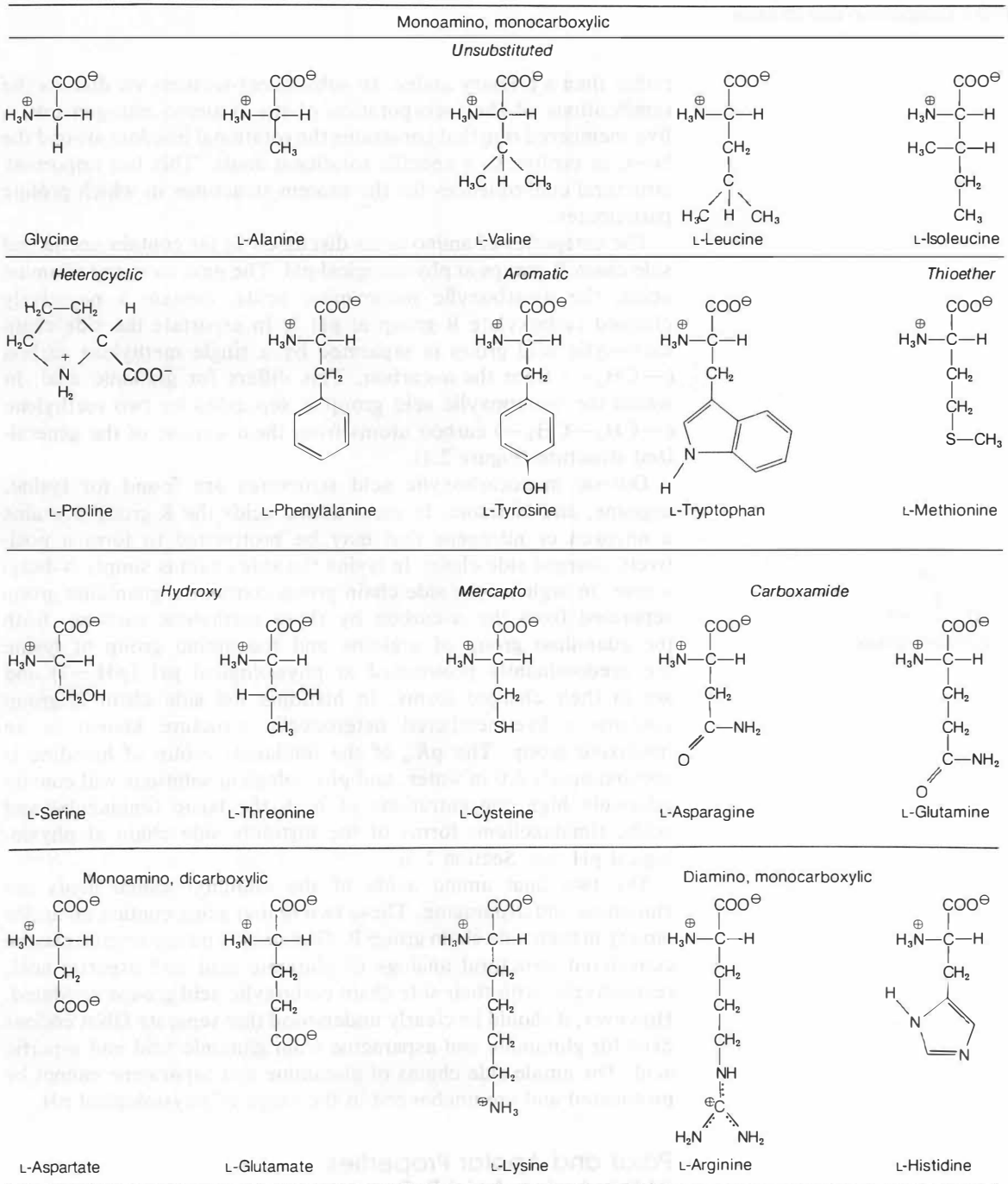
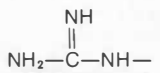


Figure 2.2
Structures of the common amino acids.

rather than a primary amine. In subsequent sections we discuss the ramifications of the incorporation of the α-amino nitrogen into a five-membered ring that constrains the rotational freedom around the N—C in proline to a specific rotational angle. This has important structural consequences for the protein structures in which proline participates.

The categories of amino acids discussed so far contain uncharged side chain R groups at physiological pH. The next category of amino acids, the dicarboxylic monoamino acids, contain a negatively charged carboxylate R group at pH 7. In aspartate the side chain carboxylic acid group is separated by a single methylene carbon ($—CH_2—$) from the α-carbon. This differs for glutamic acid, in which the γ-carboxylic acid group is separated by two methylene ($—CH_2—CH_2—$) carbon atoms from the α-carbon of the generalized structure (Figure 2.1).

Dibasic monocarboxylic acid structures are found for lysine, arginine, and histidine. In these amino acids the R group contains a nitrogen or nitrogens that may be protonated to form a positively charged side chain. In lysine the side chain is simply *N*-butyl amine. In arginine the side chain group contains a guanidino group separated from the α-carbon by three methylene carbons. Both the guanidino group of arginine and the amino group of lysine are predominantly protonated at physiological pH (pH ~7) and are in their charged forms. In histidine the side chain R group contains a five-membered heterocyclic structure known as an imidazole group. The pK_a of the imidazole group of histidine is approximately 6.0 in water, and physiological solutions will contain relatively high concentrations of both the basic (imidazole) and acidic (imidazolium) forms of the histidine side chain at physiological pH (see Section 2.3).


Guanidino group

The two final amino acids of the common amino acids are glutamine and asparagine. These two amino acids contain an amide moiety in their side chain group R. Glutamine and asparagine may be considered structural analogs of glutamic acid and aspartic acid, respectively, with their side chain carboxylic acid groups amidated. However, it should be clearly understood that separate DNA codons exist for glutamine and asparagine from glutamic acid and aspartic acid. The amide side chains of glutamine and asparagine cannot be protonated and are unchanged in the range of physiological pH.

Polar and Apolar Properties of the Amino Acid R Groups

It is important to have an appreciation for the relative hydrophobicity of the amino acid side chains in order to understand the role

played by the different amino acids in protein structure and function. The more hydrophobic (nonpolar) of the amino acids are tryptophan, phenylalanine, leucine, isoleucine, valine, tyrosine (uncharged), and proline. In globular proteins the side chain R groups from the more hydrophobic amino acids are folded into the interior of the protein molecule, away from the water of solvation. Exceptions to the above rule may occur in small regions of the total protein molecule and are usually associated with a function of the protein such as providing a hydrophobic binding site for substrate or ligand molecules to the protein. Amino acids with intermediate polarity include the smaller alkyl amino acids, glutamine, asparagine, threonine, serine, and methionine. These amino acid side chains are found both in the interior and on the solvent protein interface in significant proportions. In contrast, the amino acids containing charged R groups at pH 7 [lysine, arginine, histidine (charged), glutamate, and aspartate] are almost always found on the surface in globular proteins where the charge is stabilized by the water solvent. The rare positioning of a charged side chain into the interior of a globular protein is usually correlated with an essential structural or functional role for the "buried" charged side chain group within the nonpolar interior of the protein.

Amino Acids Are Polymerized into Peptides and Proteins

The polymerization of the 20 common amino acids into polypeptide chains within cells is catalyzed by enzymes and requires RNA and ribosomes to occur (Chapter 19). Chemically, the polymerization of amino acids into protein is a dehydration reaction. The chemical rationale of the reaction is shown in Figure 2.3. The figure shows that the α-carboxyl group of an amino acid with side chain R_1 may be covalently joined to the α-NH_2 group with side chain R_2 by the elimination of a molecule of water to form a type of amide bond known as the peptide bond. The dipeptide (two amino acid residues joined by a single peptide bond) can then form a second peptide bond through its terminal carboxylic acid group to the α-amine of a third amino acid (R_3), generating a tripeptide (Figure 2.3). Repetition of this stepwise dehydration process will generate a polypeptide or protein of specific amino acid sequence (R_1—R_2—R_3—R_4 . . . R_n). The specific amino acid sequence of a natural polypeptide is determined from the genetic information (Chapter 19). *The amino acid sequence of the polypeptide chains in a protein is known as the primary structure of the protein. It is the primary structure (amino acid sequence) that gives a protein its physical properties and causes a*

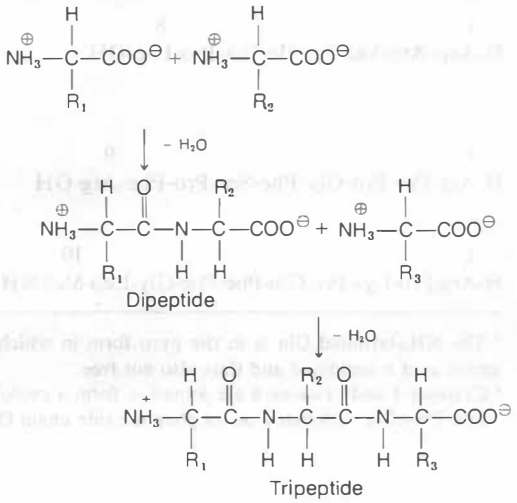


Figure 2.3
Peptide bond formation.

Table 2.1 Some Examples of Biologically Active Peptides

Amino Acid Sequence	*Name*	*Function*
1 3 pyroGlu-His-Pro(NH_2)[a]	Thyrotropic hormone releasing factor	Secreted by hypothalamus and causes pituitary gland to release thyrotropic hormone
1 9 H-Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly(NH_2)[b] S —— S	Vasopressin (antidiuretic hormone)	Secreted by pituitary gland and causes kidney to retain water from urine
1 5 H-Tyr-Gly-Gly-Phe-Met-OH	Methionine enkephalin	Opiatelike peptide found in brain that inhibits sense of pain
1 10 pyroGlu-Gly-Pro-Trp-Leu-Glu-Glu-Glu-Glu-Glu- 11 17 Ala-Tyr-Gly-Trp-Met-Asp-Phe(NH_2)[a,c] SO_3	Little gastrin (human)	Hormone secreted by mucosa cells in stomach and causes parietal cells of stomach to secrete acid
1 10 H-His-Ser-Gln-Gly-Thr-Phe-Thr-Ser-Asp-Tyr- 11 20 Ser-Lys-Tyr-Leu-Asp-Ser-Arg-Arg-Ala-Gln- 21 29 Asp-Phe-Val-Gln-Trp-Leu-Met-Asn-Thr-OH	Glucagon (bovine)	Pancreatic hormone involved in regulating glucose metabolism
1 8 H-Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-OH	Angiotensin II (horse)	Pressor or hypertensive peptide; also stimulates release of aldosterone from adrenal gland
1 9 H-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg-OH	Plasma bradykinin (bovine)	Vasodilator peptide
1 10 H-Arg-Pro-Lys-Pro-Gln-Phe-Phe-Gly-Leu-Met(NH_2)	Substance P	Neurotransmitter

[a] The NH_2-terminal Glu is in the pyro form in which its γ-COOH is covalently joined to its α-NH_2 via amide linkage; the COOH-terminal amino acid is amidated and thus also not free.

[b] Cysteine-1 and cysteine-6 are joined to form a cyclo-hexa structure within the nonapeptide.

[c] The Tyr-12 is sulfonated on its phenolic side chain OH.

polypeptide chain to fold into a unique structure and acquire its characteristic function and role.

The largest natural polypeptide chains in humans are found in protocollagen (a precursor of the structural protein collagen) and contain approximately 1,200 amino acid residues per polypeptide chain. On the other end of the spectrum there are many small peptides with less than 10 amino acids that perform important biochemical and physiological functions in human beings (Table 2.1).

Some remarks are pertinent with regard to the format used to write primary structures, as exemplified by Table 2.1. Primary structures are commonly written and numbered from their NH_2-terminal end toward their COOH-terminal end. Accordingly, for thyrotropin-releasing factor the glutamic acid residue written on the left of the sequence is the NH_2-terminal amino acid of the tripeptide and is designated amino acid residue 1 in the sequence. The proline is the COOH-terminal amino acid in the structure and is designated the third amino acid residue in the sequence. The defined direction of the polypeptide chain is from Glu → Pro (NH_2-terminal amino acid to COOH-terminal amino acid).

The standard three-symbol abbreviations for the common amino acids are used in Table 2.1. These abbreviations, given in Table 2.2, will be used almost exclusively henceforth. The abbreviations of glutamic acid (Glu) and aspartic acid (Asp) should not be confused with those for glutamine (Gln) and asparagine (Asn). In some cases the experimentalist is not able to differentiate between Gln and Glu or between Asn and Asp in a primary structure. This is because the side chain amide groups are hydrolyzed in Asn and Gln to the free carboxylic acids by the chemical procedures often utilized in the determination of the amino acid sequence of a polypeptide chain (see Section 3.4). In these cases the experimentalist will depict Gln or Glu by Glx, and Asn or Asp by Asx.

Table 2.2 The Three-Letter Symbols for the Amino Acids

Amino Acid	*Abbreviation*
Alanine	Ala
Arginine	Arg
Aspartic acid	Asp
Asparagine	Asn
Cysteine	Cys
Glycine	Gly
Glutamic acid	Glu
Glutamine	Gln
Histidine	His
Isoleucine	Ile
Leucine	Leu
Lysine	Lys
Methionine	Met
Phenylalanine	Phe
Proline	Pro
Serine	Ser
Threonine	Thr
Tryptophan	Trp
Tyrosine	Tyr
Valine	Val

Amino Acids Have an Asymmetric Center

The common amino acids with the general structure of Figure 2.1 have four substituents (R, H, COO^-, NH_3^+) covalently bonded to the α-carbon in the α-amino acid structure. A carbon atom with four different substituents arranged in a tetrahedral configuration is asymmetric and exists in two enantiomeric forms. Thus each of the amino acids (*except glycine*, in which R = H and thus two of the four substituents on the α-carbon are hydrogen) exhibits optical isomerism.

The absolute configuration for an amino acid is depicted in Figure 2.4, utilizing the Fischer projection to show the direction in space of the tetrahedrally arranged α-carbon substituents. The α-COO^- is

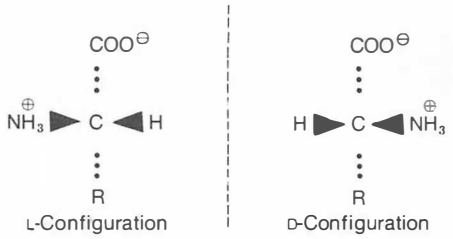


Figure 2.4
Absolute configuration of an amino acid.

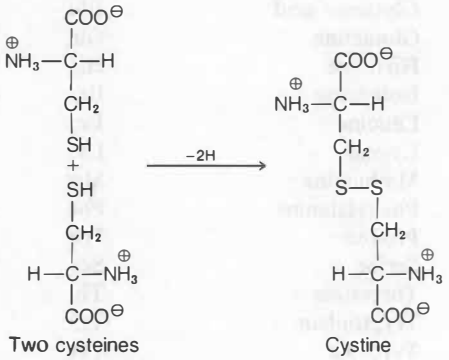


Figure 2.5.
Cystine bond formation.

directed up and back behind the plane of the page, and the side chain group R is directed down and also back behind the plane of the page. The α-H and α-NH_3^+ groups are directed toward the reader. An amino acid held in this way projects its α-NH_3^+ group either to the left or right of the α-carbon. By convention, if the α-NH_3^+ is projected to the left, the amino acid has an L absolute configuration. Its optical enantiomer, with α-NH_3^+ projected toward the right, has a designated D absolute configuration. In mammalian proteins only amino acids of L configuration are found.

As the amino acids in proteins are asymmetric, the proteins into which the amino acids are polymerized also exhibit asymmetric properties.

Derived Amino Acid Cystine

A derived amino acid found in most protein structures is cystine. Cystine is formed by the oxidation of two cysteine thiol side chain residues, which are joined to form a disulfide covalent bond (Figure 2.5). The resulting disulfide amino acid is the derived amino acid cystine. Within proteins covalent disulfide links of cystine formed from cysteines, separated from one another in the primary structure, have an important role in stabilizing the folded conformation of proteins. Cystines are formed after the free SH-containing cysteines are incorporated into the protein's primary structure and after the protein has folded.

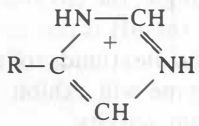
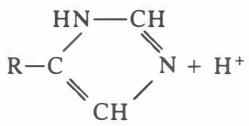
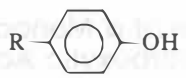
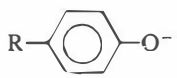
2.3 CHARGE AND CHEMICAL PROPERTIES OF AMINO ACIDS AND PROTEINS

Acid–Base Properties of the Common Amino Acids and Proteins

An understanding of proteins requires a knowledge of the ionizable side chain groups of the common amino acids. These ionizable groups common to proteins and amino acids are shown in Table 2.3.

The acid forms of the respective ionizable groups are on the left of the equilibrium sign in the table, and their respective conjugate bases are on the right of the equilibrium sign. Characteristic of the acid forms is that nitrogen-containing groups are positively charged, whereas the acid forms that contain oxygen and sulfur atoms are

Table 2.3 Characteristic pK_a Values for the Common Acid Groups in Proteins

Where Acid Group Is Found	*Acid Form*		*Base Form*	*Approximate p K_a Range for Group*
NH_2-terminal residue in peptides, lysine	$R{-}NH_3^+$ (Ammonium)	$\rightleftharpoons$	$R{-}NH_2 + H^+$ (Amine)	7.6–10.6
COOH-terminal residue in peptides, glutamate, aspartate	$R{-}COOH$ (Carboxylic acid)	$\rightleftharpoons$	$R{-}COO^- + H^+$ (Carboxylate)	3.0–5.5
Arginine	$R{-}NH{-}C(=NH_2^+){-}NH_2$ (Guanidinium)	$\rightleftharpoons$	$R{-}NH{-}C(=NH){-}NH_2 + H^+$ (Guanidino)	11.5–12.5
Cysteine	$R{-}SH$ (Thiol)	$\rightleftharpoons$	$R{-}S^- + H^+$ (Thiolate)	8.0–9.0
Histidine	 (Imidazolium)	$\rightleftharpoons$	 (Imidazole)	6.0–7.0
Tyrosine	 (Phenol)	$\rightleftharpoons$	 (Phenolate)	9.5–10.5

neutral. In contrast, the base forms of the nitrogen-containing groups are uncharged, the oxygen and sulfur-containing base forms are negatively charged.

The nitrogen atom-containing R groups (Lys, Arg) are known as the basic amino acids, as these side chains have high pK_a values and function as good bases. They are usually in their acid forms and positively charged at physiological pH. The carboxylic acid-containing side chain amino acids have relatively low pK_a values and are called the acidic amino acids. They are predominantly in their basic forms and negatively charged at physiological pH. Proteins in which the ratio (Σ Lys + Σ Arg)/(Σ Glu + Σ Asp) is greater than 1 are referred to as basic proteins. Proteins in which the above ratio is less than 1 are referred to as acidic proteins.

Table 2.4 Relationship Between the Difference of pH and Acid pK_a and the Ratio of the Concentrations of Base to Its Conjugate Acid

pH − pK_a *(Difference Between pH and p*K_a*)*	*Ratio of Base to Acid*
0	1
1	10
2	100
3	1000
−1	0.1
−2	0.01
−3	0.001

Ionized Forms of Amino Acids and Proteins; Definition of pI

From a knowledge of the pK_a for each of the ionizable acid groups in an amino acid or protein and the Henderson–Hasselbalch equation, the ionic form of the molecule can be calculated at a given pH. This is an important calculation because a change in the ionization of a protein with pH will give a molecule different functional properties at different pH values.

$$pH = pK_a + \log \frac{[base]}{[acid]}$$

Henderson–Hasselbalch equation

For example, an enzyme may require a catalytically essential histidine imidazole in its basic form for catalytic activity. If the pK_a of the catalytically essential histidine in the enzyme is 6.0, at pH 6.0 one-half of the enzyme molecules will be in the active basic (imidazole) form and one-half in the inactive acid (imidazolium) form. Accordingly, the enzyme will exhibit 50% of its potential activity. At pH 7.0, the pH is one unit above the imidazolium pK_a and the ratio of (imidazole)/(imidazolium) is 10 : 1 (Table 2.4). Based on this ratio, the enzyme will exhibit $10/(10 + 1) \times 100 = 91\%$ of its potentially maximum activity.

Titration of a Monoamino Monocarboxylic Acid

An understanding of a protein's acid and base forms and their relation to charge is more clearly understood after following the titration of the ionizable groups for the simple case of an amino acid. For example, leucine contains an α-COOH with $pK_a = 2.4$ and an

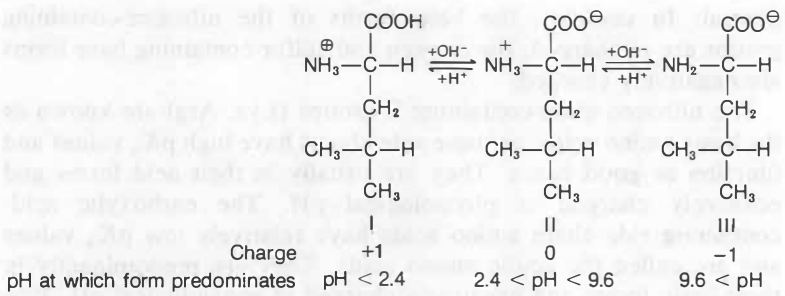


Figure 2.6
Ionic forms of leucine.

α-NH_3^+ with pK_a = 9.6. At pH 1.0 the predominant ionic form (form I) of leucine will have a formal charge of +1 and migrate toward a cathode in an electrical field. The addition of base in an amount equal to one-half of the moles of leucine present in the solution will half-titrate the α-COOH group of the leucine [i.e., $(COO^-)/(COOH) = 1$]. The pH of the solution after the addition of the 0.5 equiv of base is equal to the pK_a of the α-COOH of the leucine (Figure 2.7).

Addition of 1 equiv of base will completely titrate the α-COOH. In the predominant form (form II), the negatively charged α-COO^- and positively charged α-NH_3^+ cancel each other and the net charge on this ionic form is zero. Form II is the zwitterion form of leucine. The *zwitterion form* is that ionic form in which the positive charge from positively charged ionized groups is exactly equal to the negative charge from negatively ionized groups of the molecule. Accordingly, the net charge on a zwitterion molecule is zero, and a zwitterion molecule will not migrate toward either the cathode or anode in an electric field.

The further addition of a 0.5 mol equiv of base to the zwitterion form of leucine will half-titrate the α-NH_3^+ group. At this point in the titration, the ratio of $(\alpha\text{-}NH_2)/(\alpha\text{-}NH_3^+)$ is equal to 1, and pH = $pK_aNH_3^+$ (Figure 2.7).

The addition of a further 0.5 equiv of base (total of 2 full equiv; Figure 2.7) will completely titrate the α-NH_3^+ group. The solution pH is greater than 11.5, and the predominant species has a formal negative charge of −1 (form III).

It is useful to calculate the exact pH at which an amino acid is electrically neutral and in its zwitterion form. This pH is known as the *isoelectric pH* for the molecule, and the symbol is *pI*. The pI value is a constant of a particular compound at specific conditions of ionic strength and temperature. For simple molecules, such as amino acids, the pI is the average of the two pK_a values that form the boundaries of the zwitterion form. Leucine has only two ionizable groups, and the pI is calculated as follows:

$$pI = \frac{pK_aCOOH + pK_aNH_3^+}{2} = 6.0$$

At pH > 6.0 leucine will assume a partial negative charge that formally rises at high pH to a full negative charge of −1 [form III (Figure 2.6)]. At pH < pI leucine will have a partial positive charge until at very low pH it will have a formal charge of +1 (form I) (Figure 2.6). The partial charge at any pH can be calculated from the Henderson–Hasselbalch equation or from extrapolation from the titration curve of Figure 2.7.

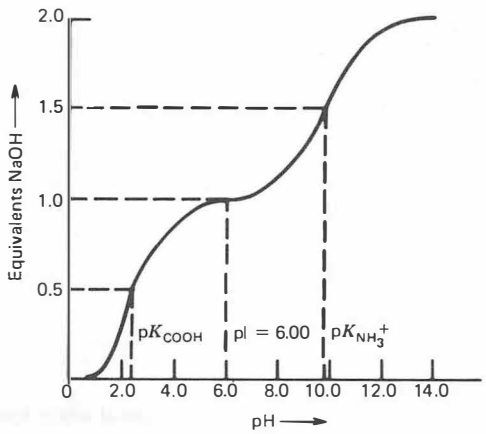


Figure 2.7
Titration curve for leucine.

Titration of a Monoamino Dicarboxylic Acid

A more complicated example of the relationship between molecular charge and pH is the example of glutamic acid. Its ionized forms and titration curve are shown in Figures 2.8 and 2.9. In glutamic acid the

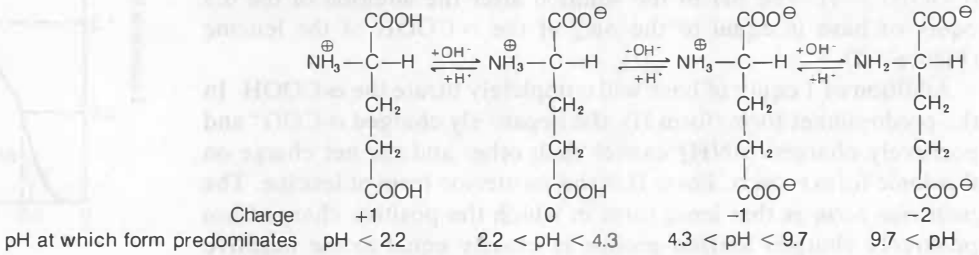


Figure 2.8
Ionic forms of glutamic acid.

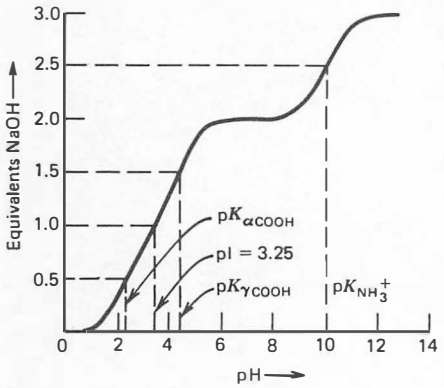


Figure 2.9
Titration curve for glutamic acid.

α-COOH pK_a = 2.2, the γ-COOH pK_a = 4.3, and the α-NH_3^+ pK_a = 9.7. The zwitterion form is generated after 1.5 equiv of base are added to the low pH form, and the isoelectric pH (pI) is calculated from the average of the two pK_a values that form the boundaries of the zwitterion form:

$$pI = \frac{2.2 + 4.3}{2} = 3.25$$

Accordingly, at values above pH 3.25 the molecule will assume a net negative charge until at high pH the molecule will have a net charge of −2. At pH < 3.25 glutamic acid will be positively charged, and at extremely low pH it will have a net positive charge of +1.

Common Charge Properties of Amino Acids and Proteins

An analysis of the charge forms present in the other common amino acids shows that the relationship found between pH and the respective pI constants for leucine and glutamate is generally true. That is, *at a solution pH less than the pI of the amino acid, the amino acid is positively charged. At a pH greater than the pI, the amino acid is negatively charged.* The degree of positive or negative charge is a function of the distance between the pH and the pI value of the amino acids and is calculable for an amino acid by the Henderson–Hasselbalch relationship.

Proteins contain multiple ionizable side chain groups and the pI value characteristic of a protein will depend on the relative concentrations of the different acid and basic R groups. As a protein contains many ionizable residues, calculation of its isoelectric pH from pK_a values would be difficult. Accordingly, the pI values for proteins are almost always experimentally measured by determining the pH value in which the protein does not move in an electrical field. The pI values found for some representative proteins are given in Table 2.5.

As with the amino acids, *at a pH greater than the pI, the protein will have a net negative charge. At a pH less than the pI, the protein will have a positive net charge. The magnitude of the net charge of a protein will increase as a function of the distance between pH and pI.* For example, human plasma albumin contains 585 amino acid residues of which there are 61 Glu, 36 Asp, 57 Lys, 24 Arg, and 16 His. The albumin pI = 4.9, at which pH the net charge is zero. At pH 7.5 the imidazolium groups of histidine have been partially titrated and albumin has a formal negative charge of −10. At pH 8.6 additional groups have been titrated to their basic form, and the formal net charge is approximately −20. At pH 11 the net charge is approximately −60. On the acid side of the pI value, at pH 3, the approximate net charge on the albumin molecule in solution is +60.

Table 2.5 pI Values for Some Representative Proteins

Protein	*pI*
Pepsin	ca.1
Human serum albumin	5.8
α_1-Lipoprotein	5.5
Fibrinogen	5.8
Hemoglobin A	7.1
Ribonuclease	7.8
Cytochrome c	10.0
Thymohistone	10.6

pH > pI, then protein charge negative
pH < pI, then protein charge positive

Separation of Amino Acids and Proteins Based on pI Values

The techniques of electrophoresis, isoelectric focusing, and ion-exchange chromatography are some of the more important techniques for the study of biological molecules based on charge.

In *electrophoresis,* an ampholyte (protein, peptide, or amino acid) in a solution buffered at a particular pH is placed in an electric field. Depending on the relationship of the buffer pH to the pI of the molecule, the molecule will either move toward the cathode (−) or the anode (+), or remain stationary (pH = pI).

An example of a classical apparatus for protein electrophoresis is shown in Figure 2.10. The apparatus consists of a U-tube in which is placed a protein solution, followed by a buffer solution carefully layered over the protein solution. The migration of the protein is observed with an optical device that measures changes in the refractive index of the solution as the protein migrates toward the anode (Figure 2.10). This apparatus historically led to the separation and operational classification of the proteins in human plasma. For the plasma protein separation, the solution is buffered at pH 8.6, which is at a pH substantially above the pI of the major plasma proteins. The proteins are negatively charged and move toward the positive

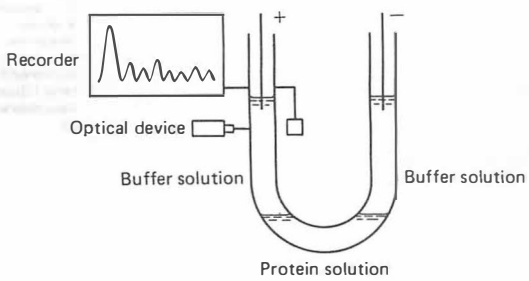


Figure 2.10
Classical electrophoresis apparatus.

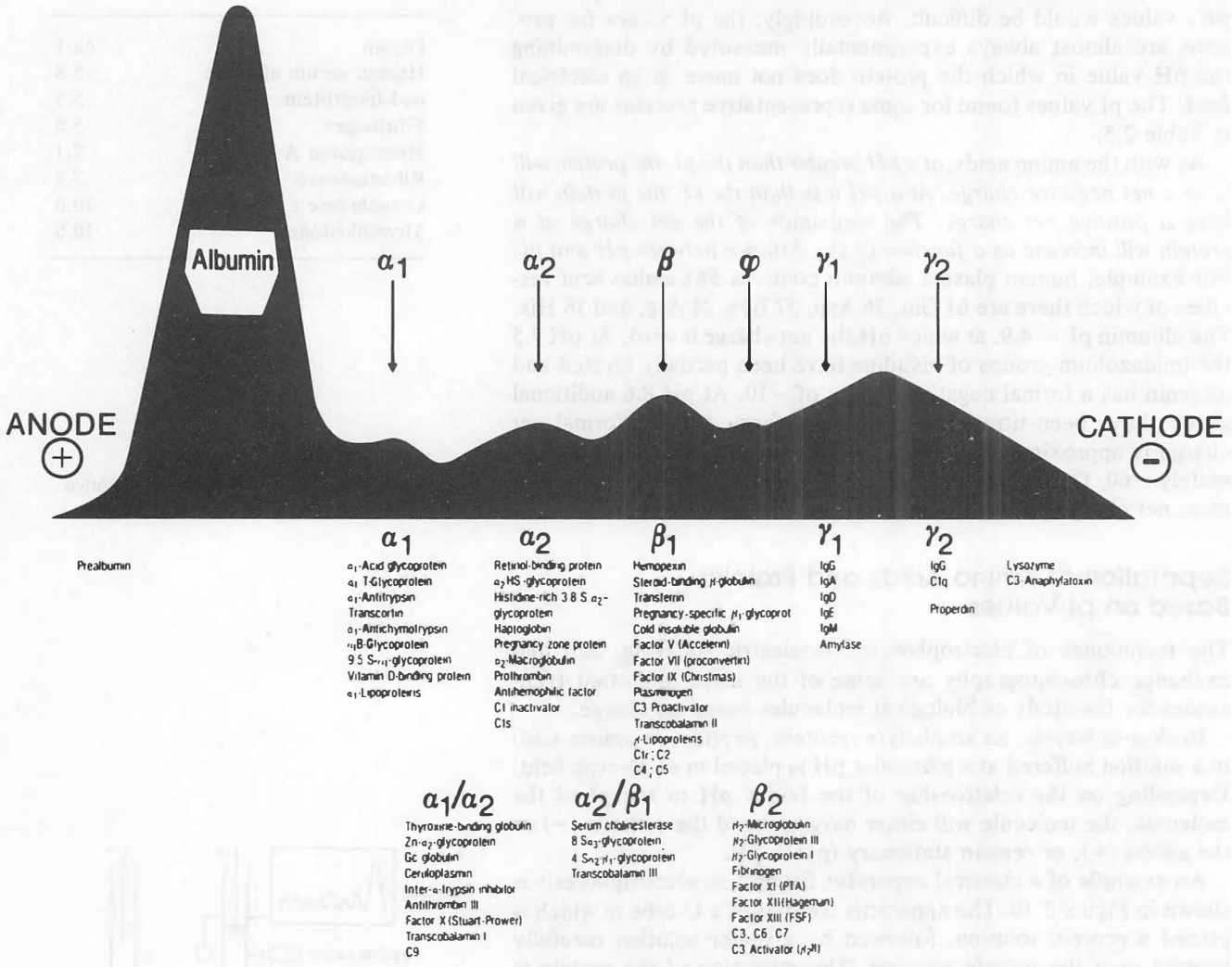


Figure 2.11
Classical (Tiselius) electrophoresis pattern for plasma proteins at pH 8.6. *The different major proteins are designated underneath the peaks. The direction of migration is from right to left with the anode (+) at left.*
Reprinted with permission from K. Heide, H. Haupt, and H. G. Schwick, in *The Plasma Proteins*, 2nd ed., Vol. III, F. W. Putnam, ed., Academic Press, New York, 1977, p. 545.

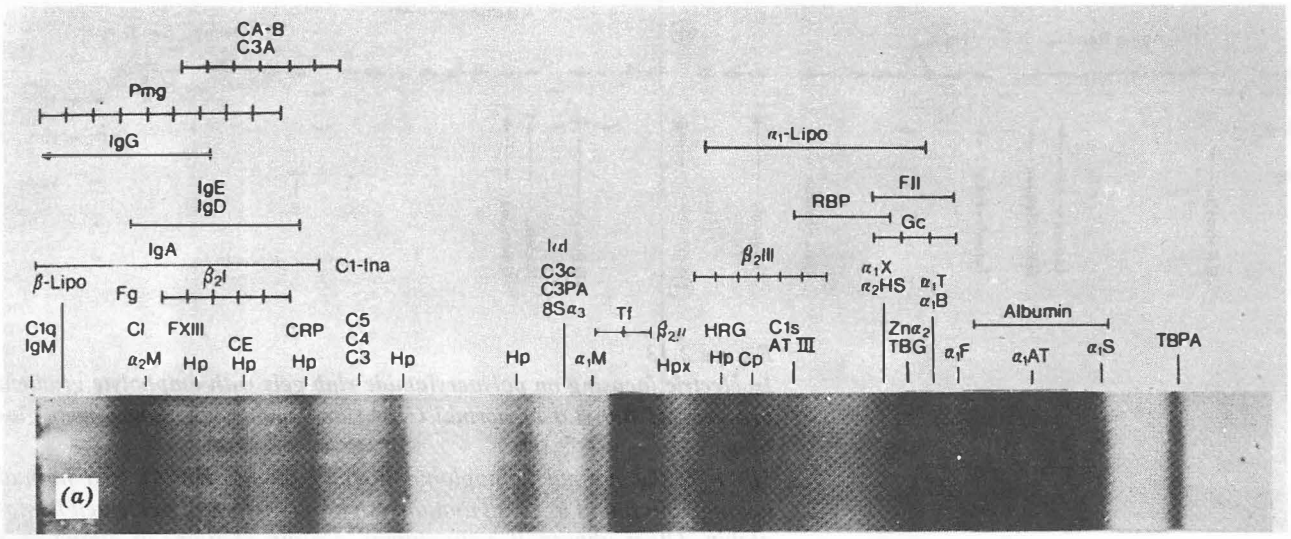
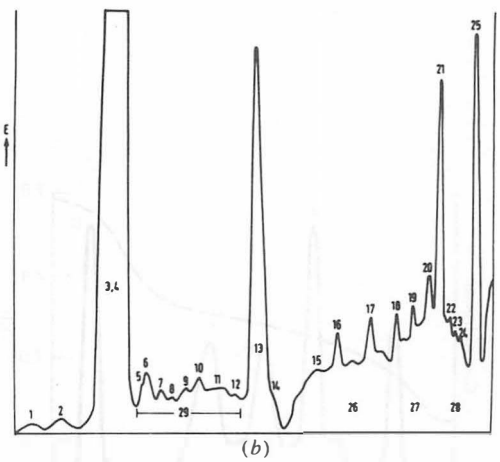


Figure 2.12

(a) Polyacrylamide gel electrophoresis of human serum proteins with anode (+) at right. (b) Densitometer trace of plasma proteins in polyacrylamide gel in an electrophoresis experiment similar to a (presented in reverse direction from (a) with anode on left).

Figure *a* from K. Heide, H. Haupt, and H. G. Schwick, in *The Plasma Proteins*, 2nd ed., Vol. III, F. W. Putnam, ed., Academic Press, New York, 1977, p. 545; Figure *b* from R. C. Allen, *J. Chromatog.*, **146**, 1 (1978). Reprinted with permission.



1, Prealbumin; 2, acid α_1-glycoprotein; 3, albumin; 4, α_1-antitrypsin; 5,7, Gc-globulins; 6, α_1 HS-glycoprotein, α_1-antichymotrypsin; 8, unknown; 9, unknown; 10, ceruloplasmin; 11, unknown; 12, hemopexin; 13, transferrin; 14, inter-α_1 trypsin inhibitor; 15, β_1-A-globulin; 16–19, haptoglobin polymers type 2–2; 20, β-glycoprotein; 21, α_2-macroglobulin; 22–24, haptoglobin polymers type 2–2; 25, β-lipoprotein; 26, IgA; 27, IgG; 28, IgM; 29, α_1-lipoprotein.

electric pole. The peaks in order of their rate of migration, related to the order of their pI values, are the albumin, the α_1-, α_2-, and β-globulins, fibrinogen, and the γ_1-, and γ_2-globulins. These peaks represent tens to hundreds of individually different plasma proteins that have a similar migration rate to the anode at pH 8.6. Their rate of migration under these experimental conditions gives an experimental value for these proteins that is widely used for purposes of their identification and classification (Figure 2.11). (A more detailed discussion of some of the significant proteins in these electrophoresis peaks is given in Chapter 3.)

More sophisticated procedures for electrophoresis utilize polymer gels, starch, or paper as an electrophoresis support. The inert supports are saturated with buffer solution, an electric field is applied across the buffered support, and the proteins migrate in the support toward a charged pole.

A common high resolution polymer support is a polyacrylamide

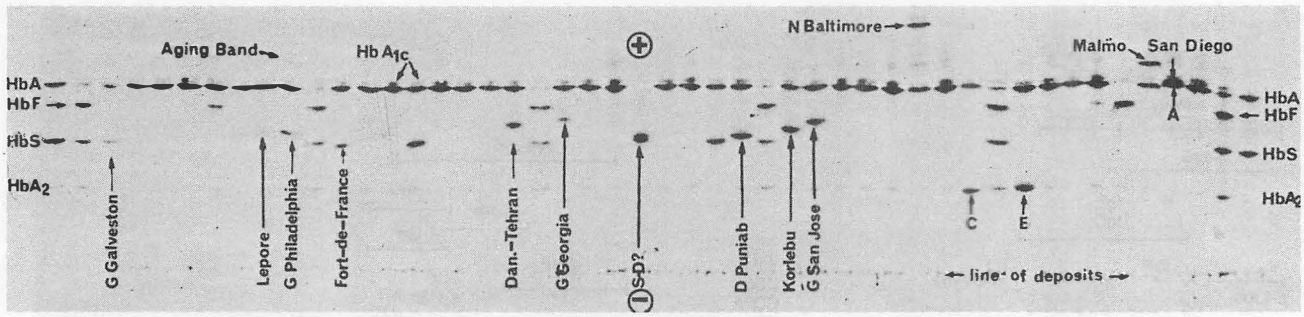


Figure 2.13
Isoelectric focusing on polyacrylamide slab gels with ampholyte gradient between pH 6.0–8.0 of normal (HbA) and abnormal human hemoglobins.
Hb F is the normal hemoglobin in fetal blood, Hb A_2 is a normal hemoglobin found in small amounts in adults, Hb S is sickle cell hemoglobin. Other abnormal hemoglobins contain changes in amino acid sequence from the normal that lead to their separation by this technique. These abnormal hemoglobins are most often named by the city from which the patient with the abnormal protein was first diagnosed.
Reprinted with permission from P. Basset, Y. Beuzard, M. C. Garel, and J. Rosa, *Blood*, **51**, 971 (1978).

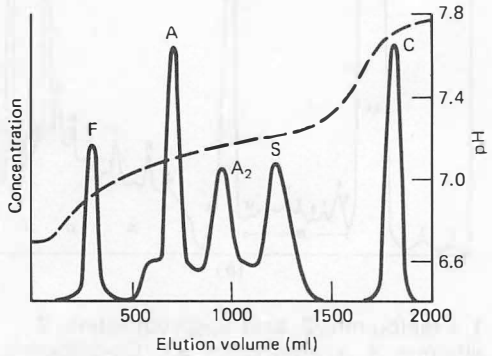


Figure 2.14
Example of ion-exchange chromatography.
Elution diagram of an artificial mixture of hemoglobins F, A, A_2, S, and C on CM-Sephadex C-50.
From A. M. Dozy and T. H. J. Huisman, *J. Chromatog.*, **40**, 62 (1969).

cross-linked gel. With the polyacrylamide gel technique, the seven peaks observed in the U-tube electrophoresis may be resolved into a greater number of distinct bands (Figure 2.12). A common *criterion for purity* of a protein is the observation of a single sharp band for a protein in a polyacrylamide gel electrophoresis experiment.

A type of electrophoresis with extremely high resolution is the technique of *isoelectric focusing;* in which mixtures of polyaminopolycarboxylic acid ampholytes with a defined range of pI's are used to establish a pH gradient across an applied electric field. A charged protein will migrate through the pH gradient in the electric field until it reaches a pH region in the gradient equal to its pI value. At this point the protein becomes stationary in the electric field and may be visualized or eluted from the column in preparative quantities (Figure 2.13). Proteins that differ as little as 0.0025 in their pI values can be separated on the appropriate pH gradient.

Separation of proteins by *ion-exchange resins* in a chromatography column is a third important technique for the separation and characterization of proteins by charge. Ion-exchange resins are prepared of insoluble materials (agarose, polyacrylamide, cellulose, glass) that contain negatively charged ligands (e.g., $—CH_2COO^-$,

$-C_3H_6SO_3^-$) or positively charged ligands covalently attached to the insoluble resin. Negatively charged resins bind cations strongly and are known as *cation-exchange resins*. Similarly, positively charged resins bind anions strongly and are referred to as *anion-exchange resins*. The degree of retardation of a protein or amino acid by a resin will depend on the magnitude of the charge on the protein at the particular pH of the experiment. Molecules of the same charge as the resin are eluted first in a single band, followed by proteins with an opposite charge to that of the resin, in an order based on the protein's charge density (Figure 2.14). In situations where it is difficult to remove a molecule from the resin because of the strength of the attractive interaction between the bound molecule and resin, systematic changes in pH or in ionic strength may be used to weaken the interaction.

For example, an increasing pH gradient in the eluent buffer through a cation-exchange resin with cationic proteins bound will reduce the difference between the solution pH and the respective molecular pI values. This decrease between pH and pI reduces the magnitude of the net charge on the proteins and thus decreases the strength of interaction between the proteins and the resin.

An increased gradient of ionic strength in the eluting buffer will also decrease the strength of charge interactions.

Some Chemical Reactions of the Amino Acids

The amino group in amino acids reacts with *ninhydrin* on heating to give a blue compound (Figure 2.15). As the reaction is stoichiometric, the amount of an amino acid present in a solution can be quantified by the light absorption of wavelength 570 nm. Proline, an imino

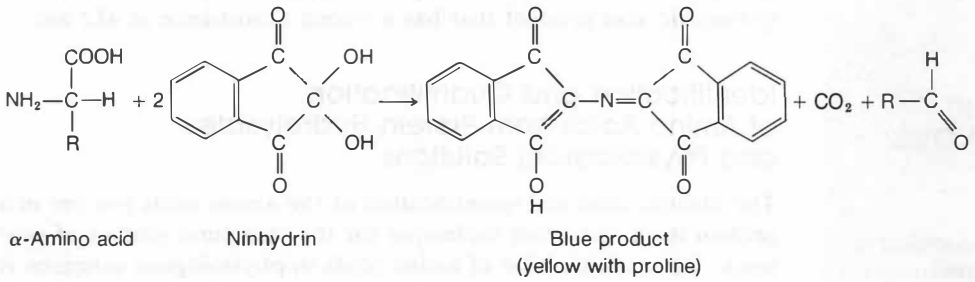
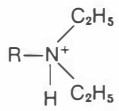


Figure 2.15
Reaction of an amino acid with ninhydrin.



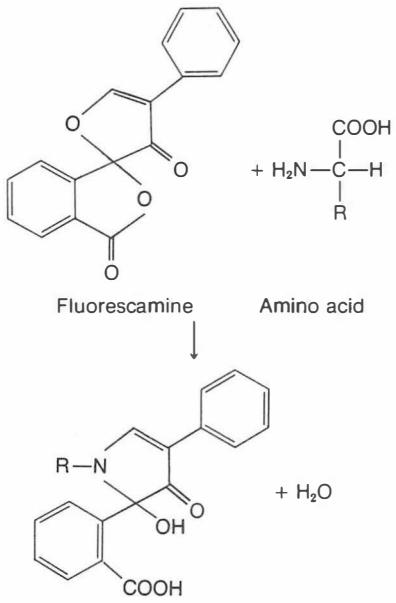


Figure 2.16
Reaction of fluorescamine with an α-amino acid to product a fluorescent product.

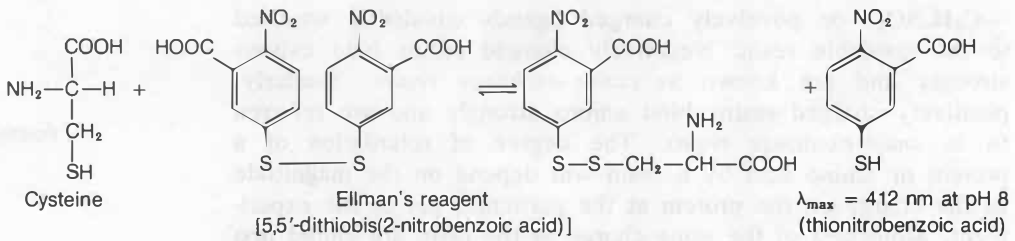


Figure 2.17
Reaction of cysteine with Ellman's reagent.

acid, also reacts with ninhydrin, but gives a yellow color with a maximum absorbance at 440 nm.

Fluorescamine reacts with amino acids in a 1:1 stoichiometry at room temperature to give a fluorescent product (Figure 2.16), the concentration of which can be determined with a spectrofluorometer. The sensitivity of the fluorescamine assay is 10 to 100 times greater than the ninhydrin assay. With current commercial instruments, the ninhydrin assay can routinely measure 3×10^{-9} mol of an amino acid, the fluorescamine procedure can measure 10^{-10} to 10^{-11} mol of an amino acid.

A goodly number of specific chemical reactions can be used to quantify specific amino acid side chains in a solution of free amino acids or in a denatured protein. The *Sakaguchi reaction* (α-naphthol and sodium hypochlorite in alkaline solution) is used to colorimetrically quantify the guanidino side chain in arginine. *Ehrlich's reagent* specifically gives a colorimetric assay for the indole group of tryptophan. The *Pauly reagent* gives a red color with the imidazole side chain of histidine and the phenol side chain of tyrosine.

Ellman's reagent [5,5'-dithiobis-(2-nitrobenzoic acid)] reacts with the cysteine thiol side chain at pH 8.0 (Figure 2.17) to give a thionitrobenzoic acid product that has a strong absorbance at 412 nm.

Identification and Quantification of Amino Acids from Protein Hydrolysates and Physiological Solutions

The identification and quantification of the amino acids present in a protein is an important technique for the structural studies of proteins. The quantification of amino acids in physiological solutions is important both for biochemical studies and for the diagnosis of disease (Clin. Corr. 2.1). The common procedure for quantitative amino acid analysis uses cation-exchange chromatography to sepa-

CLIN. CORR. 2.1 THE USE OF AMINO ACID ANALYSIS IN THE DIAGNOSIS OF DISEASE

There are a number of clinical disorders in which a high concentration of amino acids are found in plasma and/or urine. An abnormally high concentration of an amino

rate the amino acids and then the ninhydrin or fluorescamine reaction to quantitate the separated amino acids.

Cation-exchange chromatography is carried out at a pH below the pI of the common amino acids. Accordingly, the amino acids with higher pI values are retarded more strongly by the cation-exchange resin. There is also some retention of amino acids due to hydrophobic interactions with the resin, but the charge interaction is the primary cause for the separation of the amino acids by the resin. Table 2.6 gives the pI and retention time for the common amino acids obtained from a particular laboratory in chromatography over a sulfonic acid cation-exchange resin in a gradient of buffer pH values.

Table 2.6 pI Values and Elution Time of Amino Acids From a Cation-Exchange Resin[a]

Amino Acid	*Elution Time (min)*	pI[b]
Asp	9.7	2.97
Thr	11.5	6.53
Ser	12.5	5.68
Glu	14.8	3.22
Gly	19.8	5.97
Ala	21.4	6.02
Cys	24.5	5.02
Val	27.9	5.97
Met	34.3	5.75
Ile	37.1	6.02
Leu	38.0	5.98
Tyr	40.2	5.65
Phe	41.0	5.48
His	44.3	7.58
Lys	52.2	9.74
Arg	60.0	10.76

Data from J. R. Benson, *Methods Enzymol.*, **47**, 19 (1977).

[a] Amino acid analysis utilized a 0.32 × 15 cm bed of Durrum DC-5A resin. Eluent linear flow rate was 2.4 cm min^{-1}. Column temperature was 45°C from 0 to 22 min and 65°C after 22 min. Four buffers (0.2 M buffer component) at pH 3.25 (25 min), pH 4.25 (8 min), pH 5.25 (4 min), and pH 10.0 (28 min) were applied sequentially for the time duration indicated.

[b] From the lack in a good correlation between elution time and pI for some of the amino acids, the importance of hydrophobic interactions with the resin matrix to the separation (in conjunction with the ion-exchange properties of the resin) is observed.

acid in urine is called an *aminoaciduria*. *Phenylketonuria* is a metabolic defect in which patients are lacking sufficient amounts of the enzyme phenylalanine hydroxylase, which catalyzes the transformation of phenylalanine to tyrosine. As a result, large concentrations of phenylalanine, phenylpyruvate, and phenyllactate accumulate in the plasma and urine. Phenylketonuria clinically occurs in the first few weeks after birth, and if the infant is not placed on a special diet, severe mental retardation will occur. *Cystinuria* is a genetically transmitted defect in the membrane transport system for cystine and the basic amino acids (lysine, arginine, and the derived amino acid ornithine) in epithelial cells. Large amounts of these amino acids are excreted in urine. Other symptoms of this disease may arise from the formation of renal stones composed of cystine precipitated within the kidney. *Hartnup's disease* is a genetically transmitted defect in epithelial cell transport of neutral-type amino acids (monoamino, monocarboxylic acids), and high concentrations of these amino acids are found in the urine. The physical symptoms of the disease are primarily caused by a deficiency of tryptophan. These symptoms may include a pellegra-like rash (nicotinamide is partly derived from tryptophan precursors) and cerebellar ataxia (irregular and jerky muscular movements) due to the toxic effects of indole derived from the bacterial degradation of unabsorbed (untransferred) tryptophan present in large amounts in the gut. *Fanconi's syndrome* is a generalized aminoaciduria correlated with a hypophosphatemia and a high excretion of glucose.

2.4 PRIMARY STRUCTURE OF PROTEINS

Techniques Utilized to Determine the Amino Acid Sequence of a Protein

Knowledge of the primary structure (amino acid sequence) of a protein is required for an understanding of the relationship of a protein's structure to its function on a molecular level.

In the determination of primary structure, first the number of polypeptide chains in the protein must be ascertained. To begin with, the protein is denatured (see Section 2.6) and then treated with a reagent, such as dansyl chloride, that forms a covalent bond with the NH_2-terminal α-amino groups of each polypeptide chain within the protein (Figure 2.18). The tagged protein is then hydrolyzed to its constituent amino acids. Typical conditions for complete protein hydrolysis are 6 N HCl, at 110°C, for 18–36 h, in a sealed tube under vacuum. The vacuum prevents degradation of oxidation-sensitive

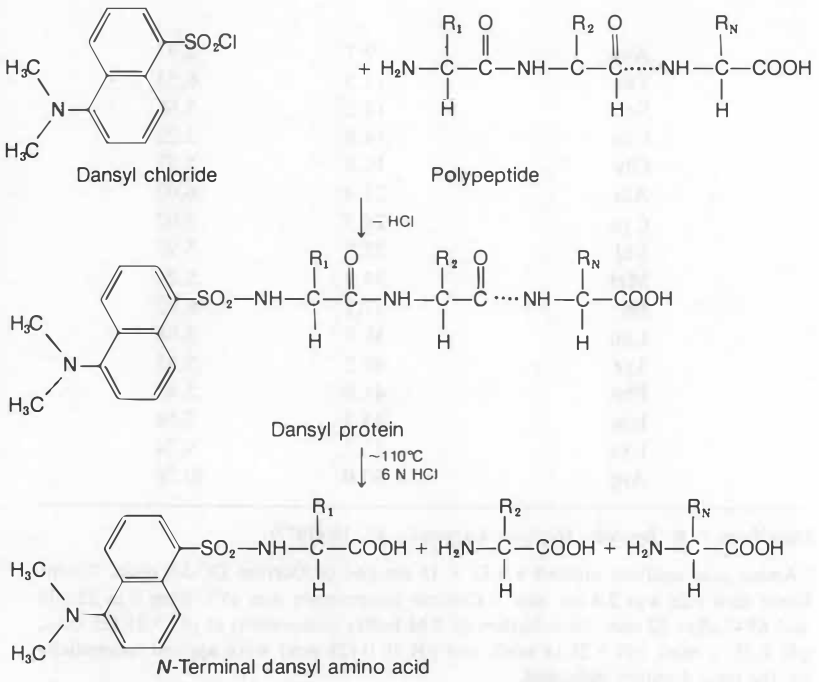


Figure 2.18
Reaction of a polypeptide with dansyl chloride.

amino acid side chains by oxygen in the air. Analysis of the amino acid hydrolysate by paper or other chromatographic procedures separates the dansyl-labeled NH_2-terminal amino acids, which are identified chromatographically against standards (Figure 2.19).

After determining the number of polypeptide chains by identification of their NH_2-terminal amino acid, the chains must be separated in order that each may be sequenced. Since the chains may be covalently joined by the disulfide bonds of cystine, these bonds will have to be broken (Figure 2.20). The individual chains are then separated by molecular exclusion chromatography (see Section 2.8), ion-exchange chromatography, and/or electrophoresis techniques.

Polypeptide chains are most commonly sequenced by the Edman reaction (Figure 2.21). In the *Edman reaction,* the polypeptide chain to be sequenced is reacted with phenylisothiocyanate, which, like the dansyl chloride, forms a covalent bond to the NH_2-terminal amino acid of the chain. However, in this derivative, acidic conditions catalyze an intramolecular cyclization that results in the cleavage of the NH_2-terminal amino acid from the polypeptide chain as a phenylthiohydantoin derivative. This NH_2-terminal amino acid derivative compound may be separated chromatographically and iden-

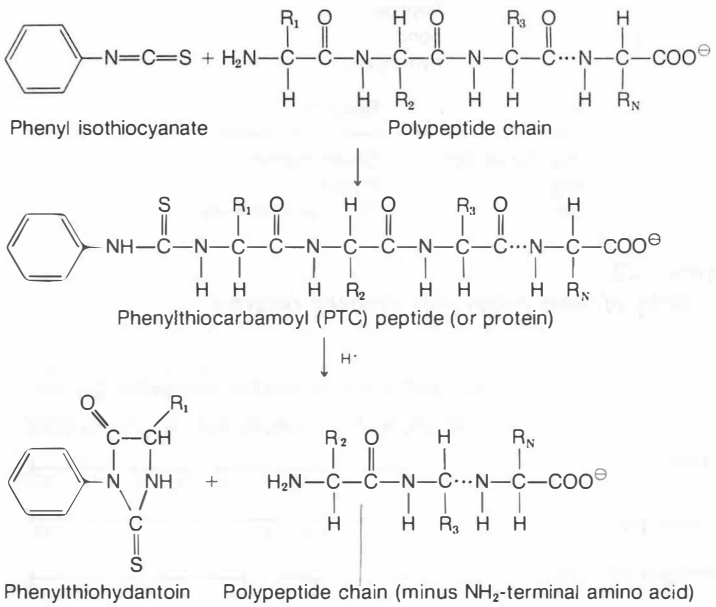


Figure 2.21
Edman reaction.

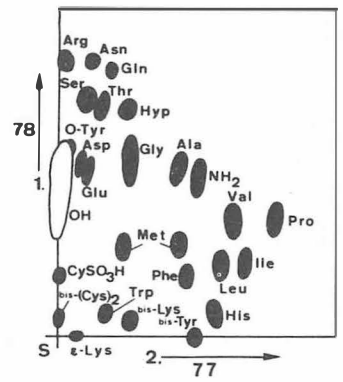


Figure 2.19
Identification of dansyl amino acids by thin-layer chromatography.
Solvent 1: water–90% formic acid (100 : 1.5 v/v); solvent 2: benzene–glacial acetic acid (9 : 1 v/v) on polyamide paper (Cheng-Chin Trading Co.).

According to D. R. Woods and K. T. Wang, *Biochem. Biophys. Acta,* **133,** 369 (1967), from A. Niederwieser in *Methods Enzymol.,* **25,** 60 (1972).

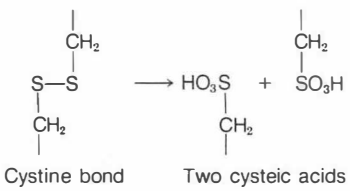
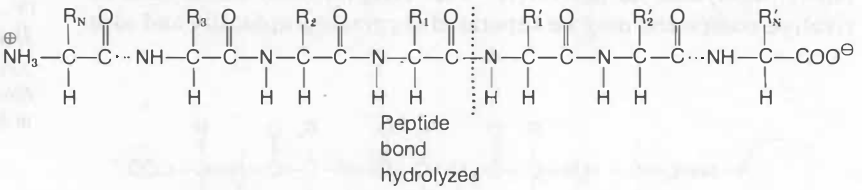


Figure 2.20
Oxidation of cystine bonds.

tified against standards. The polypeptide chain minus the NH_2-terminal amino acid is then isolated, and the Edman reaction is repeated to identify the next NH_2-terminal amino acid of the chain. This series of reactions can theoretically be repeated until the sequence of the entire polypeptide chain is determined.

The reiteration of Edman reactions under favorable conditions can be carried out for 30 or 40 amino acids into the polypeptide chain from the NH_2-terminal end. At this point in the analysis, impurities generated from incomplete reactions in the reaction series make further Edman cycles unfeasible. Since most polypeptide chains in proteins contain more than 30 or 40 amino acids, they have to be hydrolyzed into smaller fragments and sequenced in sections.

Both enzymatic and chemical methods are used to break polypeptide chains into smaller polypeptide fragments. For example, the enzymes trypsin and chymotrypsin are *proteolytic enzymes* that are commonly used for partial hydrolysis of polypeptide chains in sequencing. The enzyme trypsin preferentially catalyzes the hy-



R_1	Reagent
Phe, Tyr or Trp	Chymotrypsin
Arg, Lys	Trypsin
Met	Cyanogen bromide

Figure 2.22
Specificity of some polypeptide cleaving reagents.

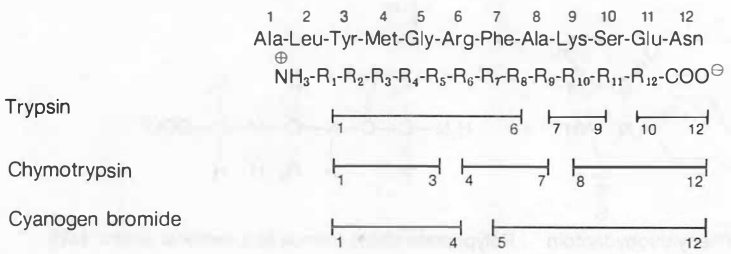


Figure 2.23
The ordering of peptide fragments from overlapping sequences produced by specific proteolysis of a peptide.

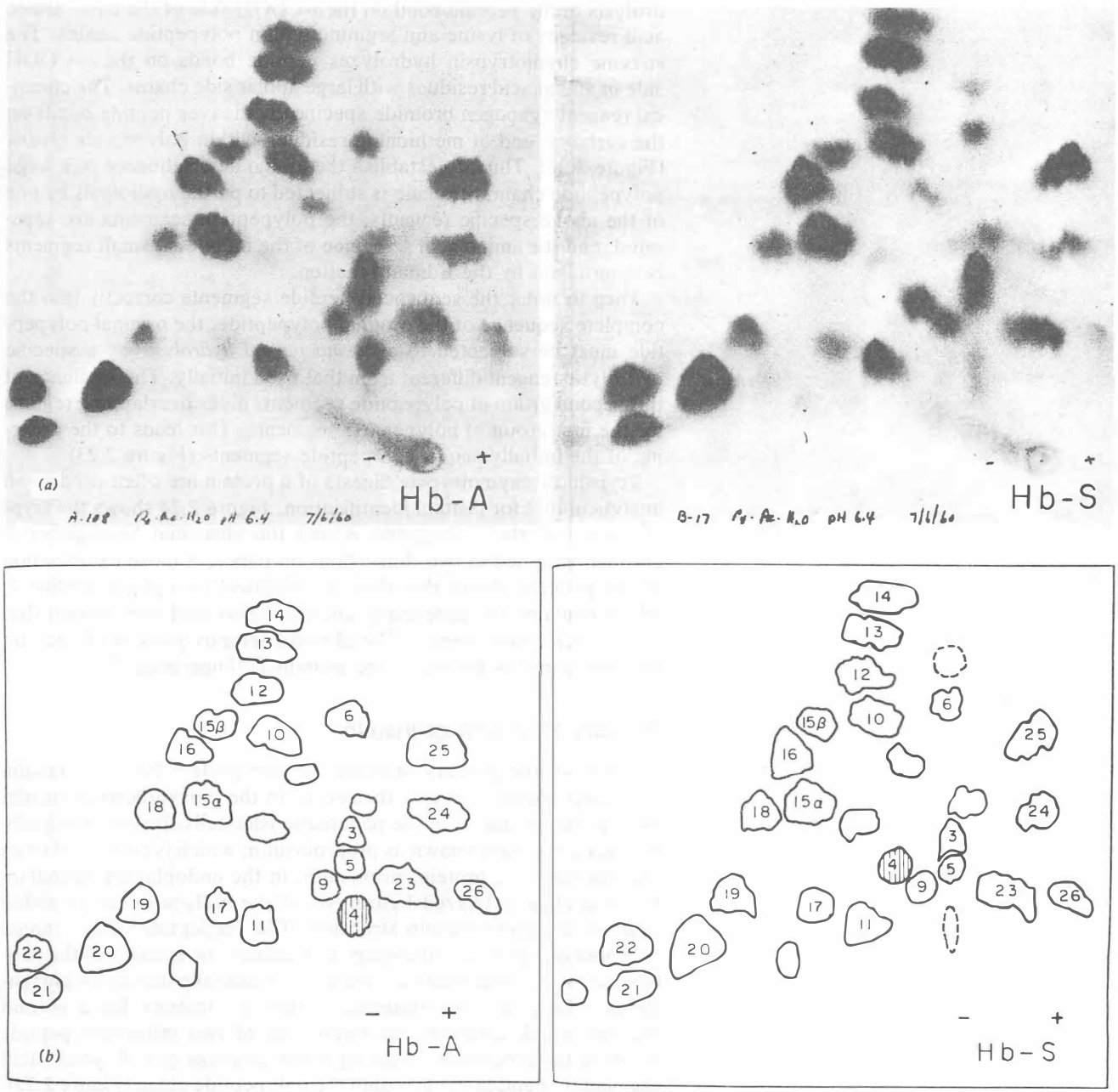


Figure 2.24
Trypsin digest of Hb A (normal) and Hb S (sickle cell hemoglobin).
(a) "Fingerprint" of digest visualized after two-dimensional chromatography. (b) Tracing of spots for peptides observed above. Dashed line tracing indicates spots that become visible only on heating. Spot 4 contains mutated amino acid that gives Hb S its sickle-cell properties; otherwise all other peptides are identical.

Reprinted with permission from C. Baglioni, *Biochim. Biophys. Acta*, **48**, 392 (1961).

drolysis of the peptide bond on the α-COOH side of the basic amino acid residues of lysine and arginine within polypeptide chains. The enzyme chymotrypsin hydrolyzes peptide bonds on the α-COOH side of amino acid residues with large apolar side chains. The chemical reagent cyanogen bromide specifically cleaves peptide bonds on the carboxyl end of methionine residues within polypeptide chains (Figure 2.22). Thus, to establish the amino acid sequence of a large polypeptide chain, the chain is subjected to partial hydrolysis by one of the above specific reagents, the polypeptide segments are separated, and the amino acid sequence of the individual small segments is determined by the Edman reaction.

Then to order the sequenced peptide segments correctly into the complete sequence of the original polypeptide, the original polypeptide must be subjected to a *second partial hydrolysis* by a specific hydrolytic reagent different from that used initially. The sequence of this second group of polypeptide segments gives overlapping regions for the first group of polypeptide segments. This leads to the ordering of the initially sequenced peptide segments (Figure 2.23).

Trypsin or chymotrypsin digests of a protein are often used as an analytical tool for protein identification. Figure 2.24 shows the trypsin digest of the hemoglobin A and the abnormal hemoglobin S chromatographed in two dimensions on paper. A close examination of the patterns shows that they are identical except for peptide 4, which contains the genetically altered amino acid substitution that causes sickle cell anemia. The chromatography pattern of such an enzymic digest is known as the protein's "fingerprint."

Primary Structure of Insulin

A study of the primary structure of the protein hormone insulin elucidates chemical events that occur in the biosynthesis of insulin from precursor forms in the pancreatic islet cells. Insulin is initially produced in a form known as preproinsulin, which is converted to an intermediate size protein, proinsulin, in the endoplasmic reticulum by an enzyme-catalyzed hydrolysis of the NH_2-terminal 23 amino acids in the preproinsulin structure. The prepeptide NH_2-terminal fragment is apparently necessary to facilitate the transfer of the peptide across the endoplasmic reticulum membrane during its biosynthesis. The proinsulin molecule is then a substrate for a second enzyme which catalyzes the hydrolysis of two intrachain peptide bonds in the proinsulin, resulting in the cleavage of a 30-amino acid segment (C-peptide) from within the polypeptide chain (Figure 2.25). The active insulin product of these proteolytic hydrolyses consists of two small polypeptide chains that are covalently held together by cystine disulfide bonds (Figure 2.26).

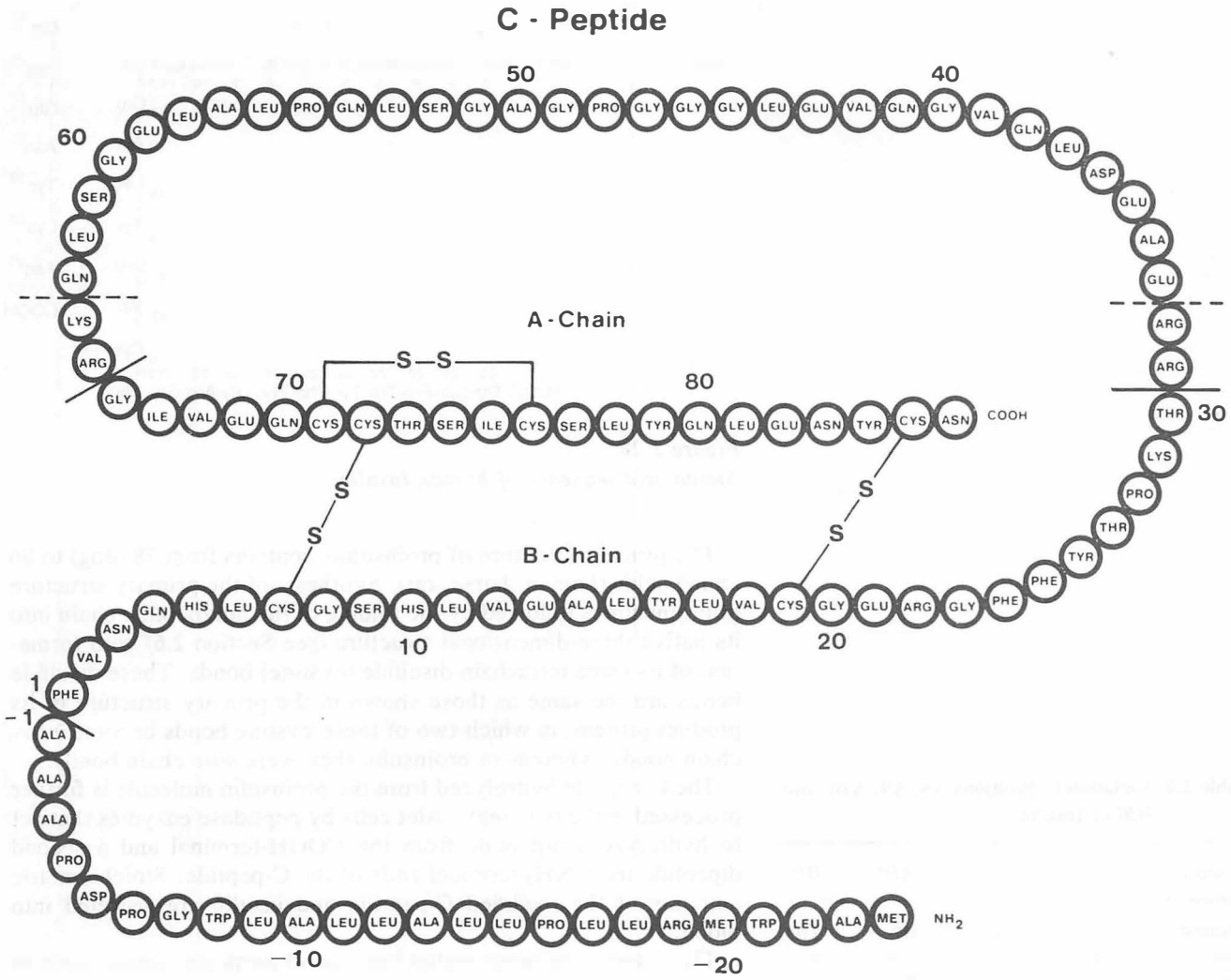


Figure 2.25
Amino acid sequence of human preproinsulin.
The prepeptide region extends from Met at position −24 to Ala at position −1, the B-peptide from Phe at position 1 to Thr at position 30, the C-peptide from Arg at position 31 to Arg at position 65, and the A-peptide from Gly at position 66 to Asn at position 86. Cystine bonds from positions 7 to 72, 19 to 85, and 71 to 76, are found in proinsulin.
From G. Bell, et al., Nature, **282,** 525 (1979).

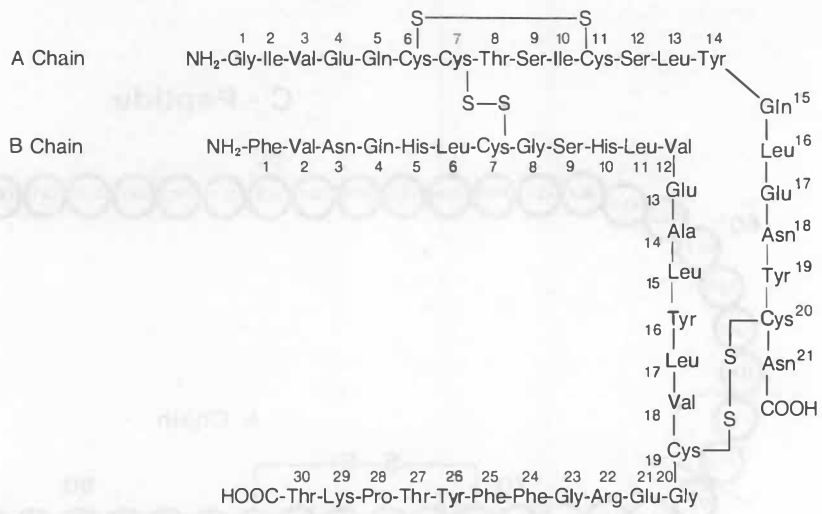


Figure 2.26
Amino acid sequence of human insulin.

The primary structure of proinsulins contains from 78 (dog) to 86 amino acids (human, horse, rat). Synthesis of the primary structure of proinsulin is followed by the folding of the polypeptide chain into its native three-dimensional structure (see Section 2.6) with formation of its three intrachain disulfide (cystine) bonds. These disulfide bonds are the same as those shown in the primary structure of its product protein, in which two of these cystine bonds become *inter*-chain bonds, whereas in proinsulin they were *intra* chain bonds.

The C-peptide hydrolyzed from the proinsulin molecule is further processed in the pancreatic islet cells by peptidase enzymes that act to hydrolyze a dipeptide from the COOH-terminal and a second dipeptide from NH_2-terminal ends of the C-peptide. Stoichiometric amounts of the modified C-peptide and insulin are secreted into blood.

The essential or nonessential function of particular amino acids in the primary structure of the active insulin is studied by the comparison of the primary structure for active insulins from different animal species. This comparison shows that residues 8, 9, and 10 of the A chain and residue 30 of the B chain may be varied without dramatically affecting the structure and corresponding physiological role of the insulin molecule (Table 2.7). The other amino acid residues of the primary structure do not vary as significantly among species, and must therefore be more critical to the structural and related functional properties of insulin.

Table 2.7 Variation in Positions A8, A9, A10, and B30 of Insulin

Species	*A8*	*A9*	*A10*	*B30*
Human	Thr	Ser	Ile	Thr
Cow	Ala	Ser	Val	Ala
Pig	Thr	Ser	Ile	Ala
Sheep	Ala	Gly	Val	Ala
Horse	Thr	Gly	Ile	Ala
Dog	Thr	Ser	Ile	Ala
Chicken[a]	His	Asn	Thr	Ala
Duck[a]	Glu	Asn	Pro	Thr

[a] Positions 1 and 2 of B chain are both Ala in chicken and duck, whereas in other species of table position 1 is Phe and position 2 is Val in B chain.

An amino acid substituted by an alternative amino acid of similar polarity (e.g., Val for Ile in position 10 of insulin) in a primary structure is designated a *conservative type* of variation (see Clin. Corr. 2.2). A *nonconservative type* of change occurs on the substitution of an amino acid by another of dramatically different polarity (see Clin. Corr. 2.3).

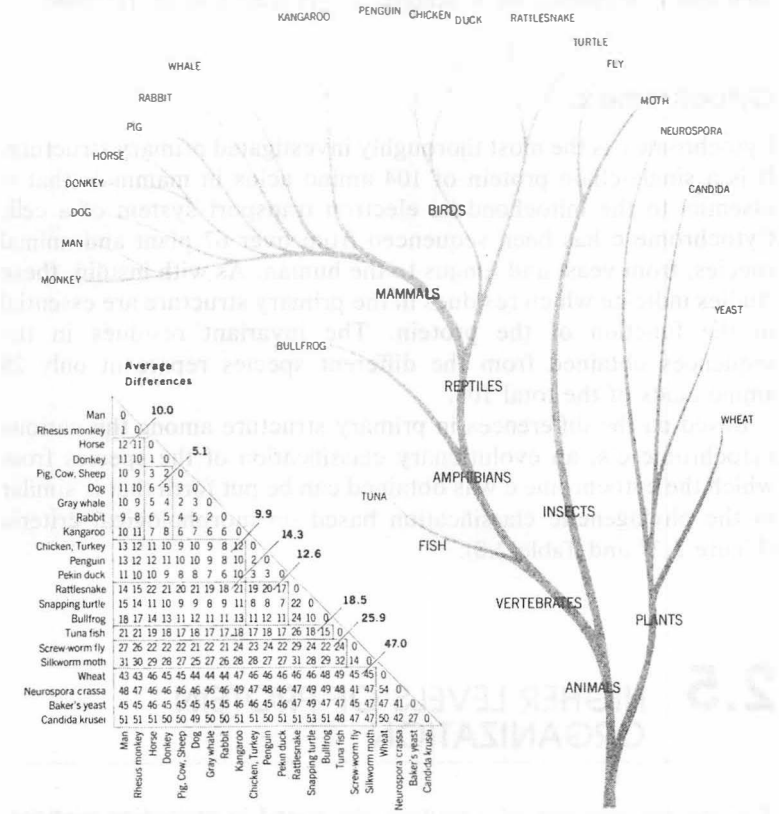


Figure 2.27
Difference in amino acid sequences among cytochrome c's from different species.
Family tree is drawn based on the calculation from number of amino acid differences.

Reprinted with permission from R. S. Dickerson and J. Geis, *The Structure and Action of Proteins*, W. A. Benjamin, Inc., Menlo Park, 1969, p. 65.

CLIN CORR. 2.2
DIFFERENCES IN THE PRIMARY STRUCTURE OF INSULINS UTILIZED IN THE TREATMENT OF DIABETES MELLITUS

Both pig (porcine) and beef (bovine) insulins are commonly used in the treatment of human diabetics. Because of the differences in amino acid sequences from the human insulin, some individuals will have an initial allergic response to the injected insulin as their immunological system recognizes the insulin as foreign, or develop an insulin resistance due to a high anti-insulin antibody titer at a later stage in treatment. The insulin resistance syndrome is more common in individuals who use insulin periodically, rather than on a continuous basis. However, the number of diabetics who have a deleterious immunological response to pig and cow insulins is small; the great majority of human diabetics can utilize the nonhuman insulins without immunological complication. The compatiblity of the cow and pig insulins in humans is due to the small number of changes and the conservative nature of the changes between the amino acid sequences of the inuslins. These changes in primary structure do not significantly perturb the structure of the insulins from that of the human insulin. Pig insulin is usually more acceptable than cow insulin in insulin-reactive individuals because it is more similar in sequence to human insulin (see Table 2.7).

CLIN. CORR. 2.3 A NONCONSERVATIVE MUTATION OCCURS IN SICKLE CELL ANEMIA

Hemoglobin S (Hb S) is a variant form of the normal adult hemoglobin in which a nonconservative substitution occurs in the sixth position of the β-polypeptide chain of the normal hemoglobin (Hb A_1) sequence. Whereas in Hb A_1 this position is taken by a glutamic acid residue, in Hb S the position is occupied by a valine. Consequently, individuals with Hb S have replaced a polar side chain group on the molecule's outside surface with a nonpolar hydrophobic side chain group (a nonconservative mutation). Through hydrophobic interactions with this nonpolar valine residue, Hb S in its deoxy conformation will polymerize with other molecules of deoxy-Hb S, leading to a precipitation of the hemoglobin within the red blood cell. The precipitation of the hemoglobin gives the red blood cell a sickle shape, an instability that results in a high rate of hemolysis and a lack of elasticity in diffusion through the small capillaries, which become clogged by the abnormal red blood cells.

Only individuals homozygous for Hb S exhibit the disease. Individuals heterozygous for Hb S contain approximately 50% Hb A_1 and 50% Hb S in their red blood cells and do not exhibit symptoms of the sickle cell anemia disease unless under extreme hypoxia.

It is of interest that individuals heterozygous in Hb S have a resistance to the malaria parasite, which spends a part of its life cycle in the red blood cell. This is a factor selecting for the Hb S gene in malarial regions of the world and the reason for the high frequency of this homozygous

Table 2.8 Evolution of Species Based on Cytochrome c Sequences

Species Compared	*Number of Residues Changed*	*Divergence of Species in Millions of Years*
Human–chicken	13	280 (assumed)
Human–tuna	21	490 (calculated)
Human–moth	31	750 (calculated)
Human–yeast	44	1180 (calculated)

Table from E. Margoliash and A. Schejter, *Adv. Protein Chem.* 21, 113 (1966).

Cytochrome c

Cytochrome c is the most thoroughly investigated primary structure. It is a single-chain protein of 104 amino acids in mammals that is essential to the mitochondrial electron transport system of a cell. Cytochrome c has been sequenced from over 67 plant and animal species, from yeast and fungus to the human. As with insulin, these studies indicate which residues in the primary structure are essential to the function of the protein. The invariant residues in the sequences obtained from the different species represent only 28 amino acids of the total 104.

Based on the differences in primary structure among the various cytochrome c's, an evolutionary classification of the species from which the cytochrome c was obtained can be put forth that is similar to the phylogenetic classification based on morphological criteria (Figure 2.27 and Table 2.8).

2.5 HIGHER LEVELS OF PROTEIN ORGANIZATION

The *primary structure* of a protein, discussed in preceding sections, refers to the covalent structure of a protein. It includes the amino acid sequence of the protein and the location of disulfide (cystine) bridges. Higher levels of protein organization refer to noncovalently generated conformational properties of the primary structure. These higher levels of protein conformation and organization are customarily defined as the secondary, tertiary, and quaternary structures of a protein. The *secondary structure* refers to the configuration of the polypeptide chain in the protein. For example, secondary structures

of polypeptide chains may form noncovalently generated configurations that are helical (i.e., α helix). The *tertiary structure* refers to the total three-dimensional structure of the protein. It includes the conformational relationships in space of the side chain groups to the polypeptide chain and the geometric relation of distant regions of the polypeptide chain with respect to each other. The *quaternary structure* refers to the configuration of the noncovalent association of discrete polypeptide subunits into a multisubunit protein when this occurs.

Proteins generally assume a unique secondary, tertiary, and quaternary conformation for their particular amino acid sequence, known as the native conformation. The folding of the primary structure into the particular structure occurs spontaneously under the influence of noncovalent forces. This unique conformation is that of the lowest Gibbs free energy kinetically accessible to the polypeptide chain(s) of the protein for the particular solvent conditions of ionic strength, pH, and temperature in which the folding process occurs. Cystine bonds are made after the folding of the polypeptide chain occurs, and act to covalently stabilize the folded conformation.

The higher levels of protein organization are individually discussed in the following sections.

lethal gene in the human genetic pool. It is shown that approximately 10% of American blacks are heterozygous in Hb S, and 0.4% of American blacks are born homozygous for Hb S and exhibit sickle cell anemia.

Secondary Structure of Proteins

The configuration for a polypeptide chain is based on the rotational angles assumed about the covalent bonds that interconnect the amino acids. These interconnecting covalent bonds are the bond between the N_α and C_α, the bond between the C_α and C′ (carbonyl carbon), and the peptide bond. The first of these is designated the phi (ϕ) bond and the second the psi (ψ) bond for an amino acid residue in a polypeptide chain.

Figure 2.28 shows that the peptide bond can be depicted by two resonance configurations. In structure I, a double bond is located between the carbonyl carbon and carbonyl oxygen, and the carbonyl carbon to nitrogen bond ($C' - N_\alpha$) is depicted as a single bond. In the electronic isomer, structure II, the carbonyl carbon to nitrogen bond is a formal double bond. Peptide bonds are a resonance hybrid of the two electron isomer representations of Figure 2.28. The carbonyl carbon to nitrogen bond ($C' - N_\alpha$) thus contains a substantial double-bond character. The consequence of the double-bond nature of the $C' - N_\alpha$ bond is that free rotation about this bond does not normally occur at physiological temperature. Accordingly, rotations occur in only two of the three bonds, (i.e. the ϕ and ψ bonds) contributed by each amino acid residue within a polypeptide chain (Figure 2.29).

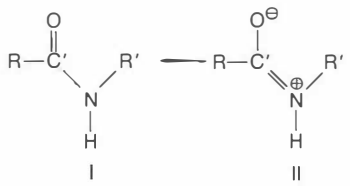


Figure 2.28
Peptide bond resonance structures.

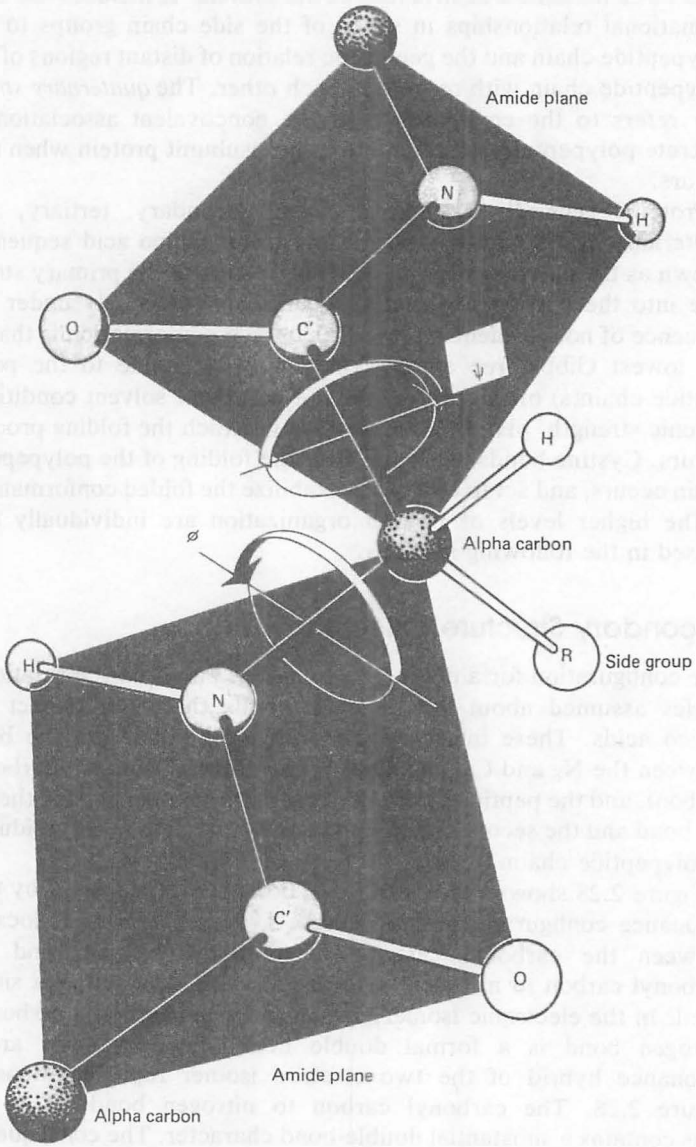


Figure 2.29
Definitions of ϕ and ψ angles.

Reprinted with permission from R. E. Dickerson and J. Geis, *The Structure and Action of Protein,* W. A. Benjamin, Inc., Menlo Park, CA. 1969, p. 25.

Regular secondary structure configurations in segments of a polypeptide chain occur when all the ϕ bond angles in that polypeptide segment are equal to each other, and all the ψ bond angles are equal. For example, when the ϕ angles $= -57°$ and the ψ angles $= -47°$, a polypeptide segment will be in an α helix configuration. When $\phi = -139°$ and $\psi = +135$, an antiparallel β structure configuration is found.

The α helix and β structure configurations for polypeptide chains are generally the most thermodynamically stable of the regular secondary structures. However, particular amino acid sequences of a primary structure in a protein may support regular configurations of the polypeptide chain other than α-helical or β structure. Thus, whereas α-helical and β structures are most commonly found, the actual configuration is dependent on the particular physical properties due to the sequence present in the polypeptide chain and the solution conditions in which the protein is dissolved. In addition, in most proteins there are significant regions of "random" structure in which the ϕ and ψ angles are not equal.

Proline will break α-helical configurations of the polypeptide chain since the pyrrolidine side chain group of proline sterically interacts with the side chain group from the amino acid preceding it in the polypeptide sequence when the preceding amino acid has a ψ angle of $-47°$ as required for the α-helical structure. This repulsive steric interaction will thereby prevent formation of

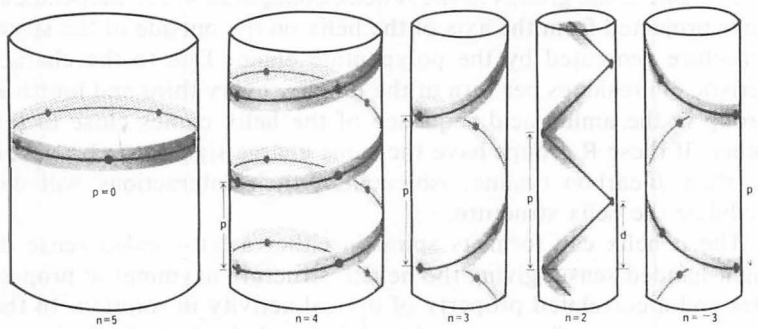


Figure 2.30
Helix parameters. Definition of pitch (p) and number of residues per turn of helix (n).
The rise per residue along the helix would be p/n (see equation in text). Each circle on line represents an α-carbon from an amino acid residue.
Reprinted with permission from R. E. Dickerson and J. Geis, *The Structure and Action of Proteins*, W. A. Benjamin, Inc., Menlo Park, 1969, p. 26.

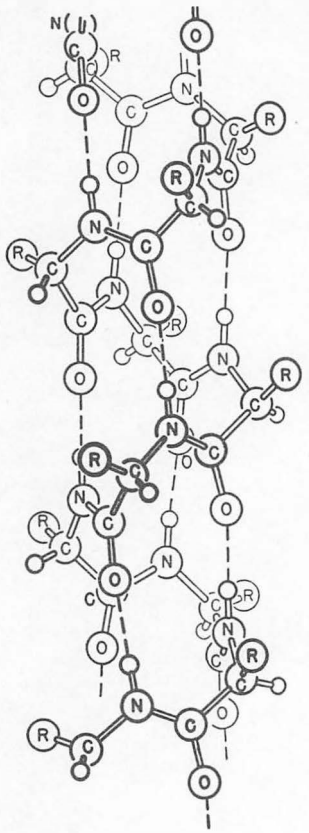


Figure 2.31
An α helix.
Reprinted with permission from L. Pauling, *The Nature of the Chemical Bond*, 3rd ed., Cornell University Press, Ithaca, N.Y., 1960.

α-helical structures in regions of a polypeptide chain in which proline is found.

As shown in Figure 2.30, the helical structures of polypeptide chains are geometrically defined by the number of amino acid residues per 360° turn of the helix (***n***) and the distance between α-carbons of adjacent amino acids measured parallel to the axis of the helix (d). The helix pitch (p), defined by the equation below, measures the distance between repeating turns of the helix on a line drawn parallel to the helix axis.

$$p = d \times n$$

α-Helical Structure

A polypeptide chain in an *α-helical configuration* is shown in Figure 2.31. Characteristic of the α-helical configuration are 3.6 amino acid residues per 360° turn ($n = 3.6$), $d = 1.5$ Å, and $p = 5.4$ Å.

The peptide bonds in the α helix are directed parallel to the axis of the helix. In this geometry each peptide forms a hydrogen bond to the peptide bond of the fourth amino acid above and the peptide bond of the fourth amino acid below in the primary structure. The distance between the hydrogen donor atom and the hydrogen acceptor atom in the α helix is 2.9 Å. Also, the donor atom, acceptor atom, and hydrogen atom in the α helix peptide hydrogen bond are colinear, that is, they lie on a straight line. This is an optimum geometry and distance for maximum hydrogen bond strength (see Section 2.6).

The side chain groups in the α helix configuration are perpendicularly projected from the axis of the helix on the outside of the spiral structure generated by the polypeptide chain. Due to the characteristic 3.6 residues per turn of the α helix, every third and fourth R group in the amino acid sequence of the helix comes close to the other. If these R groups have the same charge sign or are branched at their β-carbon (valine, isoleucine), their interactions will destabilize the helix structure.

The α helix can form its spiral in either a left-handed sense or right-handed sense, giving the helical structure asymmetric properties and a correlated property of optical activity in solution. In the structure shown, the more stable right-handed α helix is depicted.

β-Structure

A polypeptide chain in a β structure configuration is shown in Figure 2.32. In β structure, segments of a polypeptide chain are in an extended helix with $n = 2$, $d = 3.5$ Å, and $p = 6.95$ Å. The helix of one polypeptide chain or segment is hydrogen-bonded to other helices in β structure configurations. The polypeptide segments in a β structure are aligned either in a parallel or antiparallel direction to

its neighboring chains (Figure 2.32). Large numbers of polypeptide chains interhydrogen bonded in a β-type structure give a pleated sheet appearance (see Figure 2.33). In this structure the side chain groups are projected above and below the planes generated by the hydrogen-bonded polypeptide chains.

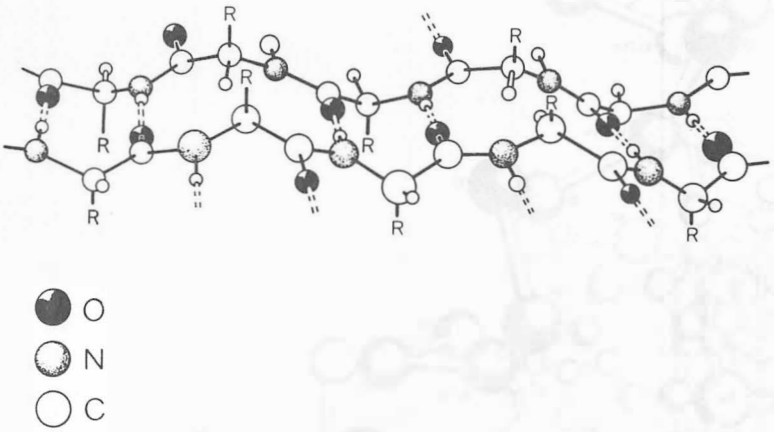


Figure 2.32
Two polypeptide chains in a β structure configuration.
Additional polypeptide chains may be added to generate more extended structure.
Reprinted with permission from A. Fersht, *Enzyme Structure and Mechanism*, Freeman, San Francisco, 1977, p. 10.

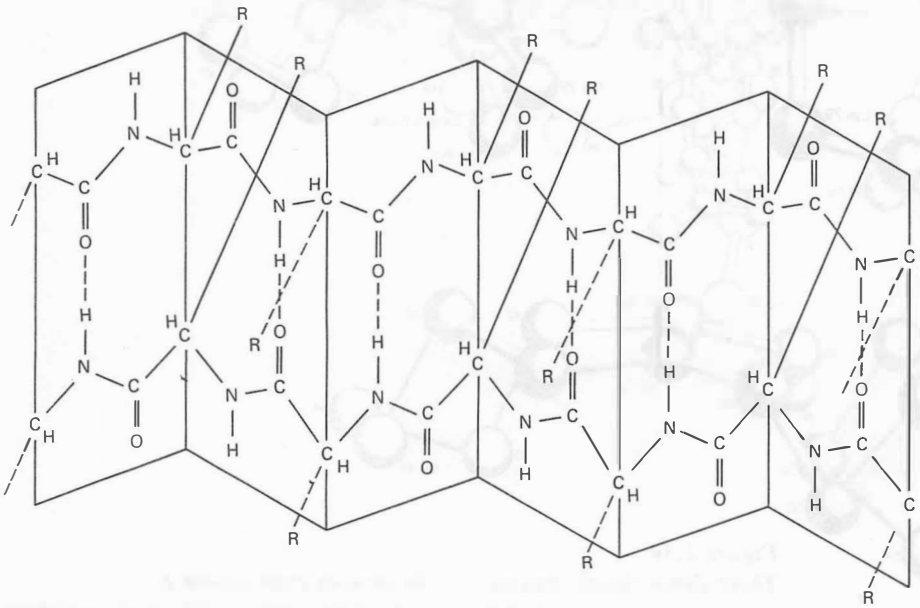


Figure 2.33
β-Pleated sheet structure between two polypeptide chains.
Additional polypeptide chains may be added above and below to generate more extended structure.

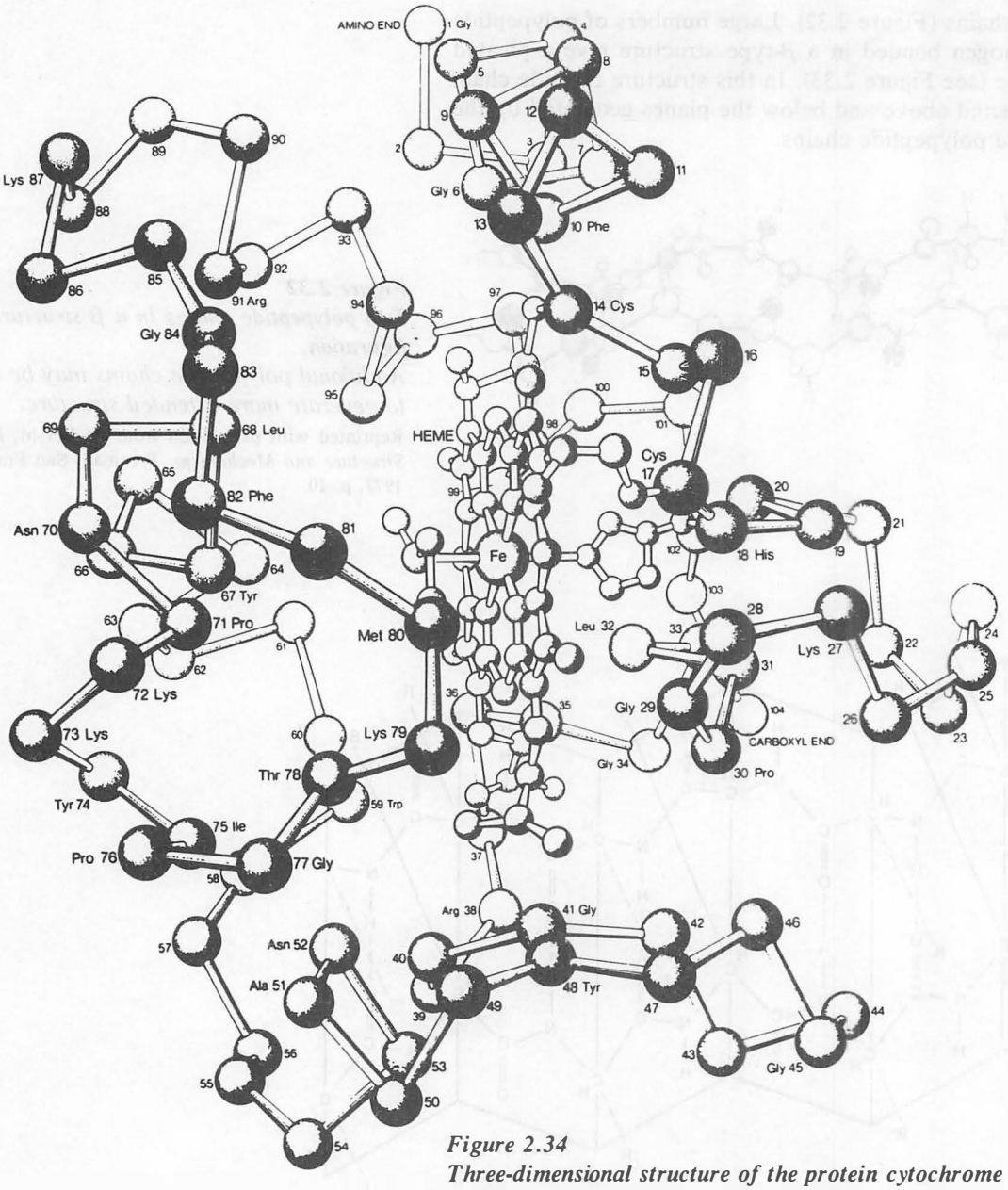


Figure 2.34

Three-dimensional structure of the protein cytochrome c.

Only α-carbons are shown for each of the amino acids in the polypeptide chain except for a methionine and histidine, which are liganded to the prosthetic group in the center of the structure. Amino acids are numbered from NH_2-terminal end (at top) to COOH-terminal end at lower right.

Reprinted with permission from R. E. Dickerson and R. Timbovich, *Enzymes*, **11**, 407 (1975).

Tertiary Structure of Proteins

The tertiary structure of a protein refers to the total three-dimensional structure of the protein, including the geometric relationship between distant segments of the primary structure and the relationship of the side chain group with respect to each other in three-dimensional space. As an example of a protein's tertiary structure, the structure for cytochrome c is shown in Figure 2.34. In this figure the protein's side chain groups are not shown in order to make the general polypeptide conformation clear. Accordingly, the ribbon diagrammatically shows the configuration of the polypeptide segments with respect to each other, with the α-carbons of the amino acid residues in the primary structure indicated by circles. Also shown are the side chain groups of Met-80 and His-18 that form coordinate bonds to an iron heme group.

The heme is a nonprotein moiety that strongly associates with the polypeptide chain of cytochrome c and is essential to the function of the protein in electron transport. A nonprotein moiety that associates with a polypeptide chain of a protein in its functional state is called a *prosthetic group*. A protein without its normal or characteristic prosthetic group is referred to as the *apoprotein*.

It is the tertiary structure that brings together the Met-80, His-18, and other residues widely separated in the primary structure to form the heme binding site of the protein.

The tertiary structure of cytochrome c conforms to the general rules of folded proteins discussed previously (Section 2.2). The hydrophobic side chains are generally into the interior of the structure, away from the water interface. Ionized amino acid side chains are found on the outside of a protein structure, where they are stabilized by water of solvation.

Within the protein structure (not shown) are buried water molecules, noncovalently associated, in specific arrangements. In general, noncovalently bound water molecules typically comprise 5 to 60% of a protein's mass by weight.

Quaternary Structure of Proteins

The quaternary structure of a protein is the arrangement or configuration of polypeptide chain units in a multisubunit protein. The subunits in a quaternary structure must be in noncovalent association. Thus the enzyme α-chymotrypsin contains three polypeptide chains covalently joined together by interchain cystine disulfide bonds into a single covalent unit. Although chymotrypsin contains three polypeptide chains, because they are covalently linked the protein is not considered to have a quaternary structure. The protein

myoglobin is composed of a single polypeptide chain and contains no quaternary structure. However, hemoglobin contains four polypeptide chains held together *noncovalently* in a specific configuration (see Chapter 3). Thus hemoglobin has a quaternary structure. The enzyme aspartate transcarbamylase (see Chapter 13) has a quaternary structure comprised of 12 polypeptide subunits. The poliovirus protein coat contains 60 polypeptide subunits, and the tobacco mosaic virus protein has 2,120 polypeptide subunits held together noncovalently in a specific structural arrangement.

2.6 FOLDING OF PROTEINS FROM RANDOMIZED TO UNIQUE STRUCTURES: PROTEIN STABILITY

The ability of a primary structure to spontaneously fold to its native secondary and tertiary conformation, without any special instructions other than the existence of noncovalent interactions, is demonstrated by experiments in which proteins are denatured without the hydrolysis of peptide bonds. These proteins, on standing, will refold to their native conformation. The experiments show that a polypeptide sequence contains sufficient physical properties to promote protein folding to the unique conformation characteristic of the protein under the correct solvent conditions and in the presence of prosthetic groups that may be a part of its structure. Quaternary structures also assemble spontaneously, after the tertiary structure of the individual polypeptide subunits are formed.

It may appear surprising that a protein always folds into a single unique conformation from all the possible a priori rotational conformations available around single bonds in the primary structure of a protein. For example, the α chain of hemoglobin contains 41 amino acids in which there are 4 to 9 single bonds per amino acid residue around which free rotation can occur. Thus there are a minimum of 4^{41} to 9^{41} possible rotational conformers for this chain. However, only a single conformation is found for the α chain. Studies on the process of protein folding indicate that certain characteristic regions of the polypeptide chain may initially associate (initiation sites). These initiation sites then promote the folding of other regions into the defined structure. Disulfide cystine bonds are formed between cysteines in the primary structure after the protein has correctly folded.

Since it is noncovalent forces that act on the primary structure to cause a protein to fold into a unique conformational structure and

then stabilize the native structure against denaturation processes, it is of importance to understand the properties of these forces. Some of the important properties of these noncovalent forces are discussed in the section that follows.

Noncovalent Forces That Lead to Protein Folding and Contribute to a Protein's Stability

Noncovalent forces are weak bonding forces of bonding strength of 1–7 kcal/mol (4–29 kJ/mol). This may be compared to the strength of covalent bonds that have a bonding strength of at least 50 kcal/mol. The noncovalent bonding forces are just higher then the average kinetic energy of molecules at 37°C (0.6 kcal/mol) (Table 2.9). However, the large number of individually weak noncovalent contacts within a protein sum to a large energy factor that is net thermodynamic force favoring protein folding.

Hydrophobic Bonding Forces

The most important of the noncovalent forces that will cause a randomized polypeptide conformation to lose rotational freedom and fold into its native structure are hydrophobic bonding forces. It is important to realize that the strength of a hydrophobic bond is not due to a high intrinsic attraction between nonpolar groups, but rather to the properties of the water solvent in which the nonpolar groups are dissolved.

A nonpolar residue dissolved in water induces in the water solvent a solvation shell in which water molecules are highly ordered. When two nonpolar groups come together on the folding of a polypeptide chain, the surface area exposed to solvent is reduced and a part of the highly ordered water in the solvation shell is released to bulk solvent. Accordingly, the entropy of the water (i.e., net disorder of the water molecules in the system) is increased. The increase in entropy (disorder) is a thermodynamically favorable process, and is the driving force causing apolar moieties to come together in aqueous solvent. A favorable free energy change of approximately 2 kcal/mol for the association of two phenylalanine side chain groups in water is due to this favorable solvent entropy gain.

Calculations show that in the folding of a randomized conformation into a regular secondary configuration such as an α helix or β structure, approximately one-third of the ordered water of solvation about the unfolded polypeptide is lost to bulk solvent. This is an approximate driving force favoring folding in a typical globular protein of 0.5–0.9 kcal/mol per peptide residue. It is calculated that an additional one-third of the original solvation shell is lost when a protein already folded into a secondary structure then folds into a

Table 2.9 Bond Strength for Typical Bonds Found in Protein Structures

Bond Type	*Bond Strength (kcal/mol)*
Covalent bonds	>50
Noncovalent bonds	0.6–7
Hydrophobic bond (i.e., 2 benzyl side chain groups of Phe)	2–3
Hydrogen bond	1–7
Ionic bond (low dielectric environment)	1–6
van der Waals	<1
Average energy of kinetic motion (37°C)	0.6

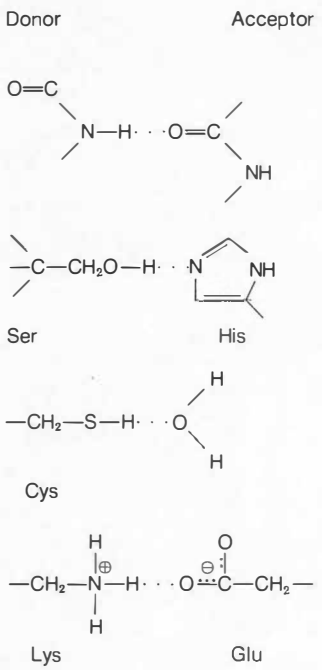


Figure 2.35
Some common hydrogen bonds found in proteins.

tertiary structure. The tertiary folding brings different segments of folded polypeptide chains into close proximity with the release of water of solvation between the polypeptide chains.

Hydrogen Bonds

A second important noncovalent force in proteins is hydrogen bonding. Hydrogen bonds are formed when a hydrogen atom covalently bonded to an electronegative atom is shared with a second electronegative atom. The atom to which the hydrogen atom is covalently bonded is designated the donor atom. The atom with which the hydrogen atom is shared is the hydrogen acceptor atom. Typical hydrogen bonds found in proteins are shown in Figure 2.35. It has been previously shown that α-helical and β structure configurations are extensively hydrogen bonded.

The strength of a hydrogen bond is to a first approximation dependent on the distance between the donor and acceptor atoms. High bonding energies occur when the donor and acceptor atoms are between 2.7 and 3.1 Å apart. Of lesser importance to bonding strength than the distance requirement, but still of some importance, is the dependence of hydrogen bond strength on geometry. Bonds of higher energy are geometrically collinear, with donor, hydrogen, and acceptor atoms lying in a straight line. The dielectric constant of the medium around the hydrogen bond may also be reflected in the bonding strength. Typical hydrogen bond strengths in proteins are 1–7 kcal/mol.

Whereas hydrogen bonds contribute to the kinetic stability of a particular protein conformation, their formation within a native protein structure is not believed to be a major driving force in the folding process. This is because peptide groups and other hydrogen-bonding groups in proteins form hydrogen bonds to the water solvent in the denatured state, and these bonds must be broken before the protein folds. The energy required to break the hydrogen bonds to water will be approximately balanced by the free energy gained by the new hydrogen bonds made between atoms in the folded protein. Therefore hydrogen bonds probably do not substantially contribute to the stability of the folded protein conformation relative to the denatured protein conformation. More explicitly, *hydrogen bonds do not have a substantial effect on the equilibrium* between the native conformation and denatured (unfolded) conformation. However, they do act in the folded conformation to place a significant energy of activation obstacle to the unfolding process that has an effect toward *decreasing the rate* at which denaturation processes occur.

Electrostatic Bonds

Electrostatic interactions between charged groups are of importance to particular protein structures and in the binding of charged ligands

and substrates to proteins. Electrostatic forces can be repulsive or attractive depending on whether the interacting charges are of the same or opposite sign. The strength of an electrostatic force (ΔE_{el}) is directly dependent on the charge (Z) for each ion, and is inversely dependent on the dielectric constant (D) of the solvent and the distance between the charge (r_{ab}).

$$\Delta E_{el} \sim \frac{Z_A \cdot Z_B \cdot \varepsilon^2}{D \cdot r_{ab}}$$

Water has a high dielectric constant (D = 80), and ionic charge interactions in water are relatively weak in comparison to electrostatic interactions in the interior of a protein, where the dielectric constant (D = 2–40) is approximately a factor of 2 to 40 times lower than in water. Consequently, the strength of an electrostatic interaction in the interior of a protein, where the dielectric constant is low, may be of significant energy. However, most charged groups of proteins are on the surface of the protein where they do not strongly interact with other charged groups from the protein due to the high dielectric constant of the water solvent, but are stabilized by hydrogen bonding and polar interactions to the water. These water interactions are the driving force leading to the placement of most ionic groups of a protein on the outside of the protein structure, where they can make energetically favorable contacts with the solvent.

Van der Waals–London Dispersion Forces

Van der Waals and London dispersion forces are a fourth type of weak noncovalent force of great importance to protein structure. This force has an attractive term (A) dependent on the 6th power of the distance between two interacting atoms (r_{ab}), and a repulsive term (B) dependent on the 12th power of r_{ab}. The A term contributes at its optimum distance an attractive force of <1 kcal/mol per atomic interaction. This attractive component is due to the inducement of complementary partial charges or dipoles in the electron density of adjacent atoms when the electron orbitals of the two atoms approach to a close distance. The repulsive component (term B) of the van der Waals force predominates at closer distances than the attractive force when the electron orbitals of the adjacent atoms begin to overlap. This type of repulsion is commonly called steric hindrance.

$$E = -\frac{A}{r_{ab}^6} + \frac{B}{r_{ab}^{12}}$$

The distance of maximum favorable interaction between two atoms is known as the van der Waals contact distance, which is

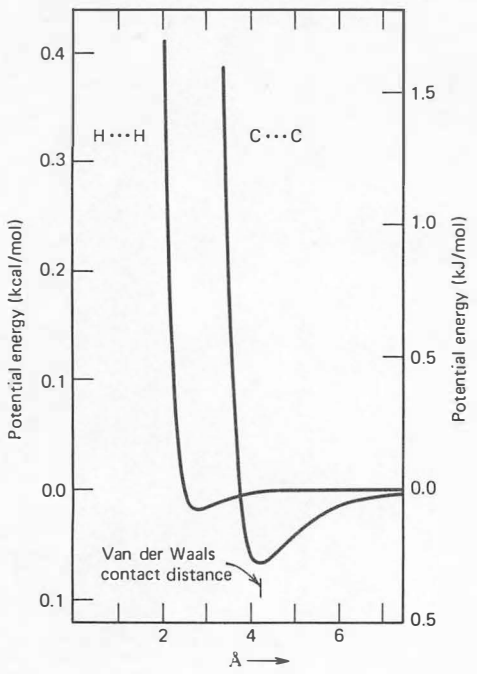


Figure 2.36
van der Waals-London dispersion interaction energies between two hydrogen atoms and two (tetrahedral) carbon atoms.
Negative energies are favorable and positive energies unfavorable.
Redrawn from A. Fersht, *Enzyme Structure and Mechanism*, W. H. Freeman, San Francisco, 1977, p. 228.

equal to the sum of the van der Waals radii for the two atoms (Figure 2.36). Some van der Waals radii for atoms commonly found in proteins are given in Table 2.10. While a London dispersion–van der Waals interaction between any two atoms in a protein is usually less than 1 kcal/mol, the total number of these weak interactions in a protein molecule is in the thousands. Thus the sum of the attractive and repulsive van der Waals–London dispersion forces are extremely important to protein folding and stability.

The van der Waals contact distances of 2.8–4.1 Å are longer than hydrogen bond distances of 2.6–3.1 Å, and at least twice as long as normal covalent bond distances of 1.0–1.6 Å between C, H, N, and O atoms. Inasmuch as the latter bonds are shorter than the van der Waals contact distance, a repulsive van der Waals force must be overcome in forming hydrogen bonds and covalent bonds between atoms. The energy to overcome van der Waals repulsive force must form a part of the energy of activation for hydrogen bond and covalent bond formation.

A special type of interaction (π-electron to π-electron) occurs when two aromatic rings approach each other with the plane of their aromatic rings overlapping (Figure 2.37). This type of interaction can result in a noncovalent attractive force of up to 6 kcal/mol.

Table 2.10 Covalent Bond Radii and van der Waals Radii for Selected Atoms

Atom	*Covalent Radius (Å)*	*van der Waals Radius (Å)* [a]
Carbon (tetrahedral)	0.77	2.0
Carbon (aromatic)	0.69 along =bond	1.70
	0.73 along —bond	
Carbon (amide)	0.72 to amide N	1.50
	0.67 to oxygen	
	0.75 to chain C	
Hydrogen	0.33	1.0
Oxygen (—O—)	0.66	1.35
Oxygen (=O)	0.57	1.35
Nitrogen (amide)	0.60 to amide C	1.45
	0.70 to hydrogen bond H	
	0.70 to chain C	
Sulfur, diagonal	1.04	1.70

SOURCE: *CRC Handbook of Biochemistry and Molecular Biology*, 3rd ed., Sect. D, Vol. II, G. D. Fasman, ed., 1976, p. 221.

[a] The van der Waals contact distance is the sum of the two van der Waals radii for the two atoms in proximity.

Denaturation of Proteins

Denaturation occurs in a protein upon the loss of its native secondary, tertiary, and quaternary structures. In the denatured state the higher structural levels of the native conformation are randomized or scrambled. However, the primary (covalent) structure is not necessarily broken in denaturation.

The loss of protein function, such as a catalytic activity in an enzyme, may occur with small modifications of the protein's native conformation that do not lead to a complete scrambling of the secondary, tertiary, and quaternary levels of conformation. Denaturation is thus correlated with the loss of a protein's function; but loss of a protein's function, which may be due to only a small conformation change, is not necessarily synonymous with denaturation.

Even though the conformational differences between denatured and native structure are substantial, the free energy difference between the denatured structure and the native structure of a protein may in some cases be as low as the free energy of a single noncovalent bond. Thus the loss of a single structurally essential hydrogen bond, or electrostatic or hydrophobic interaction can lead to denaturation of a folded structure. A change in the stability of a noncovalent bond leading to denaturation can in turn be caused by a change in pH, ionic strength, and temperature changes, which affect the strength of noncovalent bonds. The presence of prosthetic groups, cofactors, and the substrates of a protein may also affect the stability of the native configuration. These later ligands, which may have a role in the function of the protein, often also act to stabilize the native conformation of a protein.

The statement that the breaking of a single noncovalent bond in a protein can cause denaturation may appear to conflict with the observation, discussed in Section 2.4, that the amino acid sequence for a protein can often be extensively varied without loss of the native structure and related functional ability of the protein. The key to the resolution of the apparent conflict, between the extensive variability of amino acid sequence present in many proteins and the possible ease of denaturation, is the word "essential." Many noncovalent interactions in a protein are apparently not essential to the protein's overall thermodynamic stability. However, the substitution or modification of an *essential* amino acid residue of a protein that provides a critical noncovalent interaction will dramatically affect the stability of a native protein structure relative to a denatured conformation.

The concentration of a protein in vivo is under the influence of processes that both control the rate of the protein's synthesis and control the protein's rate of degradation (Figure 2.38). Therefore, an understanding of the processes that control protein degradation may

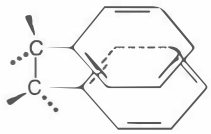


Figure 2.37
π Electron–π electron interactions between two aromatic rings.

Amino acids $\xrightarrow{\text{rate of synthesis}}$ [protein] $\xrightarrow{\text{rate of denaturation}}$ protein digest

Figure 2.38
Steady-state concentration of a protein is due to both its rate of synthesis and denaturation.

be as important as an understanding of the process of protein synthesis. It is believed that the inherent denaturation rate for a protein, in many cases, is the rate-determining step in a protein's degradation. Enzymes and cellular organelles that act to digest proteins appear to "recognize" denatured protein conformations and digest these denatured conformations at faster rates than proteins in their native conformation.

In experimental situations, denaturation of a protein can often be achieved by addition of urea or detergents (sodium dodecyl sulfate, guanidinium hydrochloride) that act to weaken hydrophobic bonding in proteins. Thus these reagents stabilize the denatured state and shift the equilibrium toward the denatured form of the protein. Addition of strong base, acid, or organic solvent, or heating to temperatures above 60°C are also common ways to cause a protein to denature.

2.7 DYNAMIC ASPECTS OF PROTEIN STRUCTURE

X-ray diffraction analysis of crystalline proteins (see Section 2.8) show that the amino acid residues within the interior of a globular protein are closely packed. The arrangement of amino acid residues appears so dense in the x-ray diffraction that in most regions of the interior the addition of a single methylene group to a side chain would prevent the side chain from fitting into the structure. However, there are observed regions of "defects" in the packing that indicate "holes" exist in the structure that give protein space for flexibility. In addition, the diffraction from some regions of a protein may not be well defined, indicating that the structure of this region, even in the crystalline state, is flexible and not well ordered. These pieces of evidence with data from nmr and fluorescence spectroscopy (see Section 2.8) show that in many proteins regions of flexibility and motion exist even into the interior. Theoretical calculations

made for particular protein structures indicate that the average atom within a protein may be fluctuating over a distance of 0.9 Å on a picosecond (10^{-9} s) time scale. Some atoms or groups of atoms will be moving smaller distances and others larger distances than this calculated average.

The possible functional significance of the internal motions within proteins is not known in most cases. However, in the proteolytic digestive enzyme trypsin, a comparison of the atomic fluctuations has been made between the active trypsin enzyme and its inactive precursor protein, trypsinogen. It is observed that regions of the trypsinogen molecule have a higher conformational disorder, correlated with a higher degree of internal atomic mobility than the active trypsin molecule. It is implied that the extent of atomic flexibility in the substrate binding region of the trypsinogen structure is a factor that decreases the ability of the protein to effectively bind and catalytically transform substrate to product. On the chemical conversion of trypsinogen to the active trypsin enzyme, the loss in atomic fluctuations at the substrate binding site facilitates the ability of the enzyme to act as a catalyst. This gives a functional significance to the relatively high degree of atomic motions for particular regions of the trypsinogen molecule in the control of the protein's activity.

2.8 METHODS FOR THE STUDY OF HIGHER LEVELS OF PROTEIN CONFORMATION AND ORGANIZATION

X-ray Diffraction Technique for the Determination of Protein Structure

The most important of the techniques for the study of a protein's secondary, tertiary, and quaternary structure is x-ray diffraction. The technique requires a protein in the crystalline form, although valuable information has also been obtained with fiber diffraction of noncrystalline materials that have a high degree of order. This latter type of x-ray diffraction has been especially important in the determination of fibrous protein structures such as that of collagen (Chapter 3).

Normally soluble globular proteins are crystallized from aqueous

solutions of high salt concentration or on addition of a miscible organic reagent. Protein molecules associate into a crystal lattice as a subunit of the geometric building units of crystal structure known as the unit cell. The crystal also includes a significant amount of water of solvation that is noncovalently incorporated into the crystalline structure.

A beam of x-rays is diffracted by the electrons around each of the atomic nuclei in the crystal, with an intensity proportional to the number of electrons around the nucleus. With the highest resolution now available to protein structure determinations, the electron diffraction from C, N, O, and S atoms can be observed. However, the diffraction from hydrogen atoms is not clearly observed due to the low number of electrons around the hydrogen nuclei.

The intensity of the diffracted x-ray beam is observed with a photographic plate or collected by an electronic device. Whereas the intensity and the angles of deflection are discernible by such data collection, the phases of the diffracted beam are not directly determined. The solution of phase for the diffracted ray commonly requires the placement of heavy atoms in the protein molecule. If heavy atoms of high electron density (such as iodine, mercury, or lead) are placed in the molecule, their position in the unit cell of the crystal can be easily located. Usually at least two different heavy atom crystalline structures, in which the crystalline structure of the protein *is unperturbed by the heavy atom modification,* must be studied before the phase problem is solved.

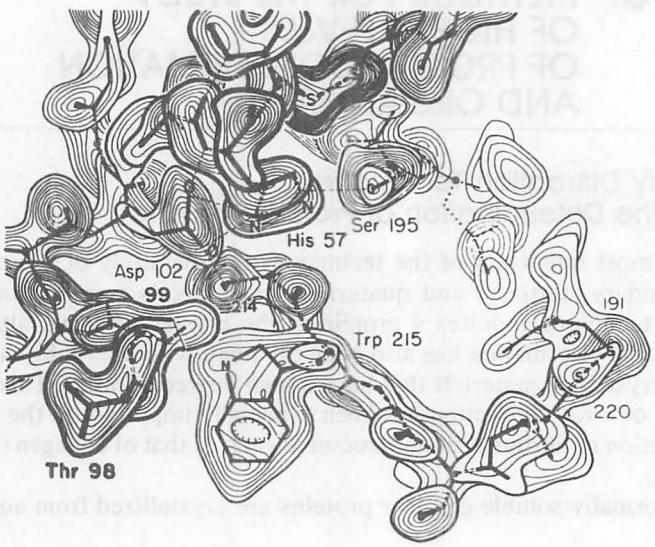


Figure 2.39
Electron-density map at 1.9 Å resolution of active site region of proenzyme (trypsinogen) form of trypsin.
The catalytically active residues in trypsin (Asp-102, His-57, and Ser-195) are superposed on the map.

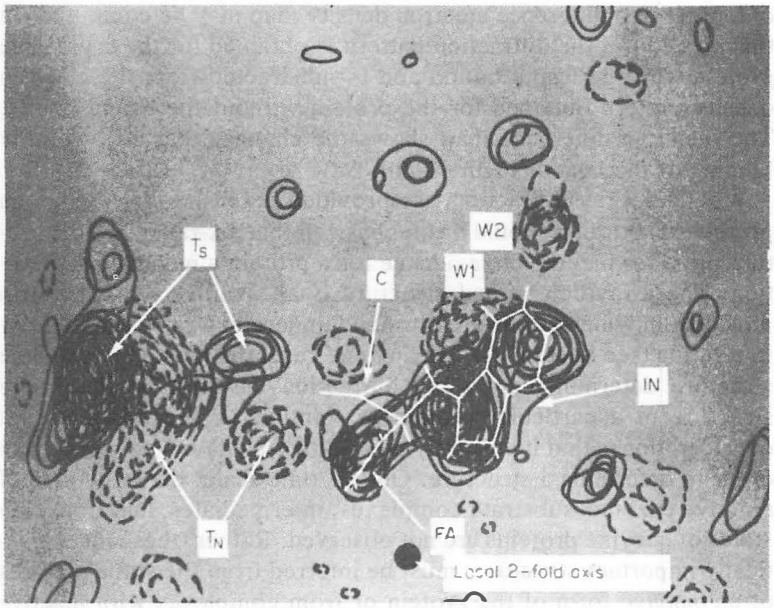


Figure 2.40
Difference electron density map between the structures of the enzyme α-chymotrypsin with substrate analog (N-formyl-L-tryptophan) bound to the enzyme active site and native α-chymotrypsin, at 2.5 Å resolution.
The smooth contours (positive density) represent electron density present in the enzyme : substrate complex and not in the native enzyme; the dashed contours (negative density) represent density present in native enzyme and not in the enzyme : substrate complex of the enzyme. Density arising from N-formyl-L-tryptophan is at IN (indole) and FA (formylamido) with structure superimposed. Negative density at W1 and W2 is due to water molecules displaced by substrate binding. The negative density T_n is the position of the carboxy terminal of Tyr-146 in the native enzyme, which moves to position T_s in the enzyme : substrate complex.
Reprinted with permission from T. A. Steitz, R. Henderson, and D. M. Blow, *J. Mol. Biol.*, **46,** 337 (1969).

Once the phases of the diffraction intensities are known from the heavy atom crystal structure, three-dimensional electron density maps can be calculated. Initially a few hundred reflections are obtained to construct a low resolution electron density map at about 6 Å. For example, in one of the first protein crystallographic structures, 400 reflections were utilized to obtain a 6-Å map of the protein myoglobin. At this level of resolution it is possible to clearly locate the molecule within the unit cell of the crystal and study the overall packing of the subunits in a protein with a quaternary structure. A trace of the polypeptide chain of an individual protein molecule may be made with difficulty. However, utilizing the low resolution structure as a base, further reflections may be used to obtain higher resolution maps. For myoglobin, whereas 400 reflections were utilized to obtain the 6-Å map, 10,000 reflections were needed for a 2-Å map, and 17,000 reflections for an extremely high resolution 1.4-Å map. A two-dimensional slice through a three-dimensional electron density map is shown in Figure 2.39. The known primary structure of the protein is fitted to the electron density pattern (Figure 2.39). The process of aligning a protein's primary structure to the electron density pattern until the best fit is obtained is known as refinement.

In order to observe changes in a protein's conformation that may occur on the binding of an inhibitor, activator, or substrate molecule

of a protein, a difference electron density map may be computed. In this procedure, the diffraction pattern is obtained for the crystalline protein with the ligand bound and is substracted from the electron density pattern obtained for the protein without the ligand bound. The resulting difference map shows the changes that occur on the binding of the ligand (Figure 2.40).

Whereas x-ray diffraction has provided extensive knowledge on protein structure, it should be emphasized that a x-ray derived structure provides incomplete evidence for a protein's mechanism of action. The x-ray-determined structure is an average structure of a molecule in which atoms are normally undergoing rapid fluctuations in solution (see Section 2.7). In any one case, the average crystalline structure determined by x-ray diffraction may not be the active structure of a particular protein in solution. A second important consideration is that it currently takes at least a day to collect data in order to determine a structure. On this time scale, the structures of reactive enzyme–substrate complexes, intermediates, and transition states of enzyme proteins are not observed. Rather, these mechanistically important structures must be inferred from the static pictures of an *inactive* form of the protein or from complexes with inactive analogs of the normally reactive substrates of the protein.

Spectroscopy

Ultraviolet Light Spectroscopy

The side chain groups of tyrosine, phenylalanine, tryptophan, and cystine, as well as the peptide bonds in proteins, can absorb ultraviolet light. The efficiency of light energy absorption for each of these different types of absorbing chromophores is related to a molar extinction coefficient (ε), which has a characteristic value for each type of chromophoric group.

A typical protein ultraviolet spectrum is shown in Figure 2.41. The absorbance between 260 and 300 nm is due to phenylalanine R groups, tyrosine R groups, and tryptophan R groups. The molar extinction coefficients for these chromophoric amino acids are plotted in Figure 2.42. When the tyrosine side chain is ionized at high pH (tyrosine R group $pK_a \simeq 10$), the absorbance for tyrosine is shifted to higher wavelength (red-shifted) and its molar absorptivity is increased (Figure 2.42).

The peptide bond absorbs in the far-ultraviolet (180–230 nm). A peptide bond in a helix configuration interacts with the electrons of other peptide bonds above and below it in the spiral configuration to create an *exciton system* in which electrons are delocalized. The result is a shift of the absorption maximum from that of an isolated peptide bond to either a lower or higher wavelength (Figure 2.43).

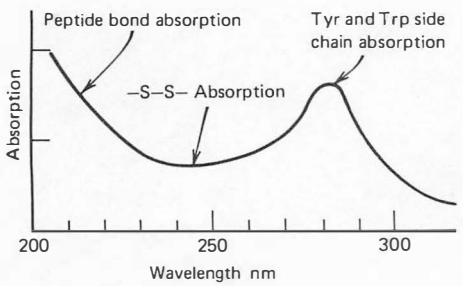


Figure 2.41
Ultraviolet absorption spectrum of the globular protein α-chymotrypsin.

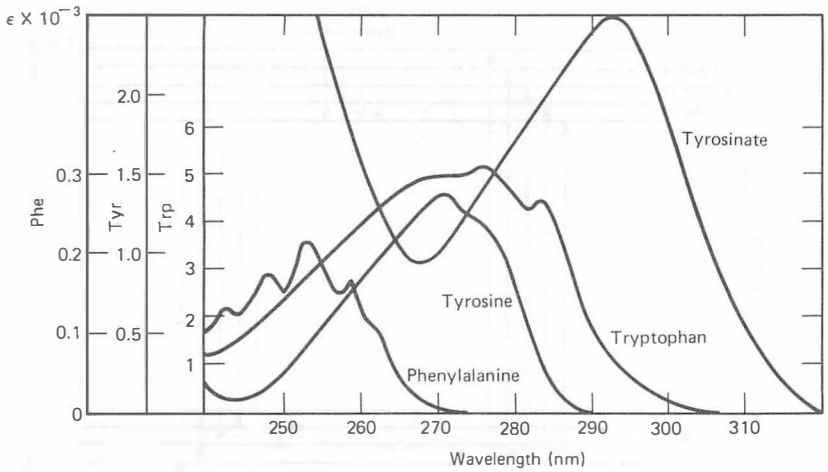


Figure 2.42
Ultraviolet absorption for the aromatic chromophores in Phe, Tyr, Trp, and tyrosinate.
Note differences in extinction coefficients on left axis for the different chromophores.
From A. d'Albis and W. B. Gratzer, in *Companion to Biochemistry*. A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton, eds., Longmans, London, 1974, p. 170.

Thus ultraviolet spectroscopy can be used to study changes in a protein's secondary and tertiary structure. As a protein is denatured (helix unfolded), differences are observed in the absorption characteristics of the peptide bonds between 180 and 230 nm due to the disruption of the exciton system. In addition, the absorption maximum for an aromatic chromophore appears at a lower wavelength in an aqueous environment than in a nonpolar environment.

The molar absorbancy of a chromophoric substrate or ligand will often change on binding to a protein. This change in the binding molecule's extinction coefficient can be used to measure its binding constant. Changes in chromophore extinction coefficients during enzyme catalysis of a chemical reaction can often be utilized to obtain the kinetic parameters for the reaction.

Fluorescence Spectroscopy

The energy of an excited electron produced by light absorption can be lost by a variety of mechanisms. Most commonly the excitation energy is dissipated as thermal energy in a collision process. In some chromophores the excitation energy is dissipated by fluorescence.

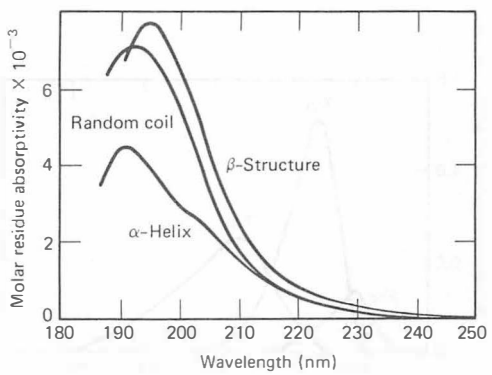


Figure 2.43
Ultraviolet absorption of the peptide bonds of a polypeptide chain in α-helical, random-coil, and antiparallel structure configurations.
From A. d'Albis and W. B. Gratzer, in *Companion to Biochemistry*, A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton, eds., Longmans, London, 1970, p. 175.

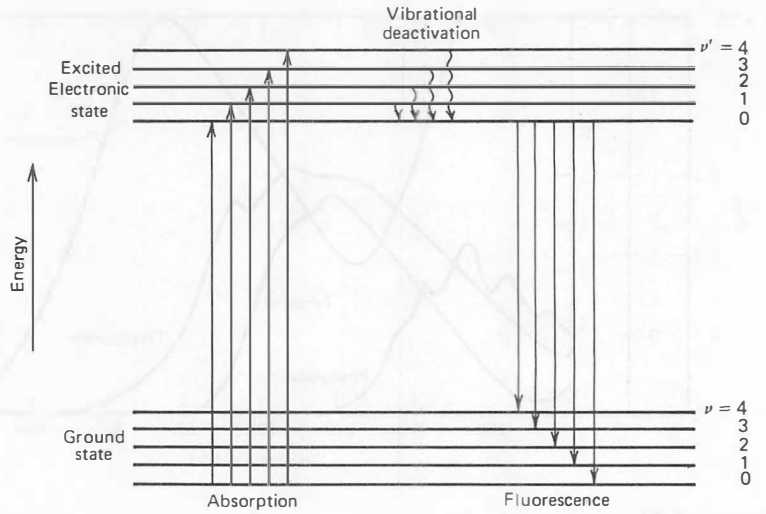


Figure 2.44
Absorption and fluorescence electronic transitions.
Excitation is from the zero vibrational level in the ground state to various higher vibrational levels in the excited state. Fluorescence is from the zero vibrational level in the excited electronic state to various vibrational levels in the ground state.

From A. d'Albis, and W. B. Gratzer, in *Companion to Biochemistry*, A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton, eds., Longmans, London, 1970, p. 166.

The fluorescent emission is always at a longer wavelength of light (lower energy) than the absorption wavelength of the fluorophore. This is because vibrational energy levels formed in the excited electron state during the excitation event are lost during the time it takes the fluorescent event to occur (Figure 2.44).

In the presence of a second chromophore that can absorb light energy at the wavelength of a fluorophore's emission, the fluorescence that is normally emitted may not be observed. Rather, the fluorescence energy can be transferred to the second molecule. The acceptor molecule, in turn, can then either emit its own characteristic fluorescence or lose its excitation energy by an alternative process. If the acceptor molecule loses its excitation energy by a nonfluorescent process, it is acting as a *quencher* of the donor molecule's fluorescence. The efficiency of excitation transfer is dependent on the distance and the orientation between donor and acceptor molecules as well as the degree of overlap between the emission wavelengths of the donor molecule and the absorption wavelengths characteristic of the acceptor molecule.

The fluorescence emission spectra for phenylalanine, tyrosine, and tryptophan side chains are shown in Figure 2.45. A comparison of their emission and absorption spectra (Figure 2.42) show that the emission wavelengths for phenylalanine overlaps with the absorption wavelengths for tyrosine. In turn, the emission wavelengths for tyrosine overlap with the absorption wavelengths for tryptophan. Because of the overlap in emission and absorption wavelengths, primarily only the tryptophan fluorescence is observed in proteins that contain all three of these types of amino acids.

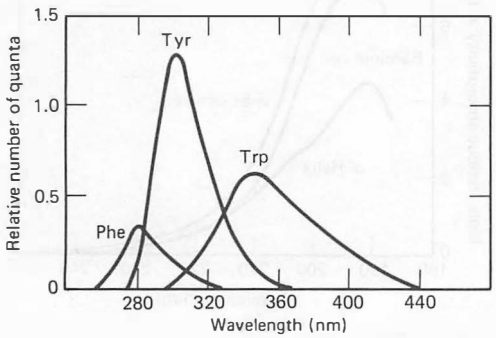


Figure 2.45
Characteristic fluorescence of aromatic groups in proteins.

From A. d'Albis and W. B. Gratzer, in *Companion to Biochemistry*, A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton, eds. Longmans, London, 1970, p. 478.

Excitation energy transfers occur over distances up to 80 Å, which are typical diameter distances in folded globular proteins. When a protein is denatured the distances between donor and acceptor groups become greater. The increased distance between donor and acceptor groups decreases the efficiency of energy transfer. Accordingly, an increase in the intrinsic fluorescence of the tyrosines and/or phenylalanines to the protein's emission spectrum will be observed with denaturation.

Since excitation transfer processes in proteins are distance- and orientation-dependent, the fluorescence yield is dependent on the conformation of the protein. As such, fluorescence is a highly sensitive tool with which to study protein conformation and changes in a protein's conformation related to its function.

Common prosthetic groups in enzyme proteins such as NADH and pyridoxal phosphate (see chapters on metabolism) are fluorophores. The changes in fluorescence yields from enzymes that contain these prosthetic moieties can be used to follow the chemical reactions catalyzed by the enzymes.

Optical Rotary Dispersion and Circular Dichroism Spectroscopy

Optical rotation is caused by differences in the *refractive index* encountered by the clockwise and counterclockwise vector components of a beam of polarized light in a solution containing an asymmetric solute. *Circular dichroism* is caused by differences in the *light absorption* between the clockwise and counterclockwise component vectors of a beam of polarized light. In proteins the aromatic amino acids of asymmetric configuration give an optical rotation and circular dichroism. Also the polypeptide chains in regular helical configuration will form either a right-handed or left-handed direction spiral configuration. These two spiral configurations are not superimposable, and they generate a significant optical rotation and circular dichroism.

Circular dichroism spectra for different configurations of the polypeptide chain are shown in Figure 2.46. Due to the differences, circular dichroism is a sensitive assay for the amount and type of secondary structure in a protein.

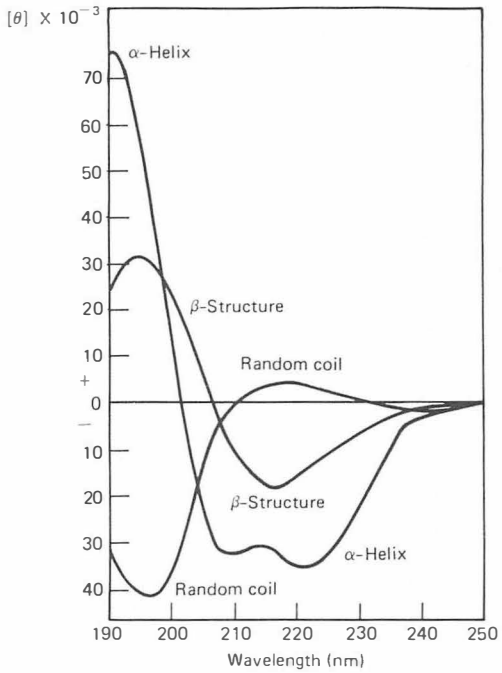


Figure 2.46
Circular dichroism spectra for polypeptide chains in α-helical, β structure, and random-coil configurations.

From A. d'Albis and W. B. Gratzer, in *Companion to Biochemistry*, A. T. Bull, J. R. Lagmado, J. O. Thomas, and K. F. Tipton, eds., Longmans, London, 1970, p. 190.

Nuclear Magnetic Resonance (nmr)

The nucleus of the atomic isotopes ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P have a characteristic spin number of one-half and in a magnetic field will absorb energy. The resonance frequency of the absorption, the pattern exhibited, and the time constants for relaxation of the excited spin state are dependent on the environment and structure into which the excited nucleus is incorporated. Accordingly, the nmr spectrum

of nuclei in a protein structure will give structural and functional information on the nucleus and its microenvironment within the protein.

Separation and Characterization of Proteins Based on Molecular Weight or Size

Ultracentrifugation; Definition of Svedberg Coefficient (*s*)

A protein subjected to centrifugal force will move in the direction of the force at a velocity dependent on the protein's mass. The rate of movement can be measured with the appropriate optical system, and from the measured rate of movement the sedimentation coefficient (s) calculated in Svedberg units (units of 10^{-13} s). In the equation,

$$s = \frac{\gamma}{\omega^2 r}$$

which can be used to calculate a sedimentation coefficient for a molecule, γ is the measured velocity of protein movement, ω the angular velocity of the centrifuge rotor, and r the distance from the center of the tube in which the protein is placed to the center of rotation. Sedimentation coefficients between 1 and 200 Svedberg units (S) have been found for proteins (Table 2.11).

Equations have been derived to relate the sedimentation coefficient to the molecular weight for a protein. One of the more simple equations is shown below, in which R is the gas constant, T the

Table 2.11 Svedberg Coefficients for Some Plasma Proteins of Different Molecular Weights

Protein	$S_{20,w} \times 10^{-13}$ *cm/s dyne*[a]	*Mol Wt*
Lysozyme	2.19	15,000–16,000
Albumin	4.6	69,000
Immunoglobulin G	6.6–7.2	153,000
Fibrinogen	7.63	341,000
C1q (factor of complement)	11.1	410,000
α_2-Macroglobulin	19.6	820,000
Immunoglobulin M	18–20	1,000,000
Factor VIII of blood coagulation	23.7	1,120,000

SOURCE: *CRC Handbook of Biochemistry and Molecular Biology*, 3rd ed., Sect. A, Vol. II, G. D. Fasman, ed., 1976, p. 242.

[a] $S_{20,w} \times 10^{-13}$ is sedimentation coefficient in Svedberg units, referred to water at 20°C, and extrapolated to zero concentration of protein.

temperature, s the sedimentation coefficient, D the diffusion coefficient of the protein, $\bar{\nu}$ the partial specific volume of the protein, and ρ the density of the solvent. The quantities D and $\bar{\nu}$ must be measured in independent experiments. In addition, *the equation assumes a spheroid geometry for the protein.* In view of the fact that the assumption of a spheroid geometry may not be true for any particular case, and independent measurements of D and $\bar{\nu}$ are difficult, the sedimentation coefficient for a molecule is often only reported. The magnitude of the protein's sedimentation coefficient will give a relative value that can be used in a generally qualitative way to characterize a protein's molecular weight.

$$\text{Molecular weight} = \frac{RTs}{D(1 - \bar{\nu}\rho)}$$

Molecular Exclusion Chromatography

A porous gel in the form of small insoluble beads is commonly used to separate proteins by size in column chromatography. Small protein molecules can penetrate the pores of the gel and will have a larger solvent volume through which to travel in the column than large proteins, which are sterically excluded from the pores. Accordingly, a protein mixture will be separated by size, the larger proteins eluted first, followed by the smaller proteins, which are retarded by their accessibility to a larger solvent volume (Figure 2.47).

As with ultracentrifugation, an assumption must be made as to the geometry of the unknown protein, and nonspheroid proteins will give anomalous molecular weights when compared to standardizing proteins of spherical conformation.

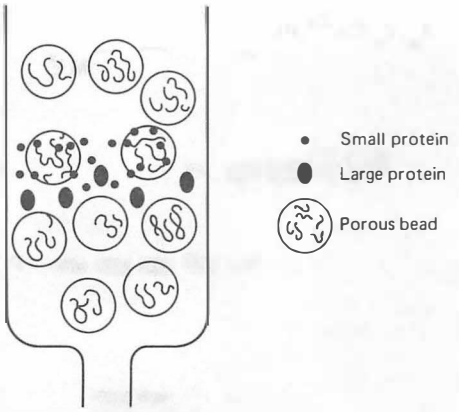


Figure 2.47
Molecular exclusion chromatography.
The small protein can enter the porous gel particles and will be retarded on the column with respect to the larger protein that cannot enter the porous gel particles.

Polyacrylamide Gel Electrophoresis in the Presence of a Detergent

If a charged detergent is added to an electrophoresis buffer and a protein is electrophoresed on a sieving support, a separation of proteins occurs based on protein size and not charge. A detergent commonly used in protein electrophoresis based on size is sodium dodecyl sulfate (SDS). The dodecyl sulfates are amphiphilic 12-carbon alkyl sulfate molecules that act to denature the protein and form a charged micelle about the denatured molecule. The inherent charge of the native protein is obliterated by the charged micelle layer of SDS, and each protein–SDS-solubilized aggregate has an identical charge per unit volume due to the charge characteristics of the SDS micelle. The negatively charged micelle particles will move through an electrophoresis gel toward the anode (+ pole). A common gel for SDS electrophoreses is cross-linked polyacrylamide. In

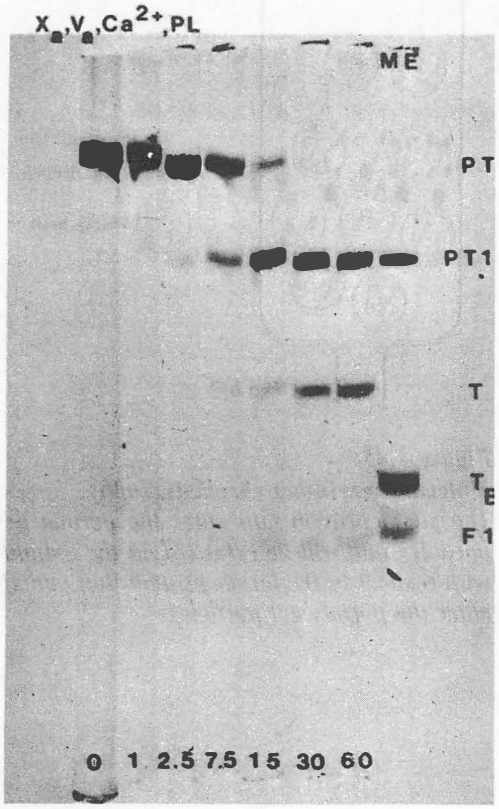


Figure 2.48
Example of sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis.
Time course of activation of the coagulation enzyme thrombin as followed by SDS-electrophoresis: PT band, prothrombin; PT1, prethrombin (intermediate activation product); T, active thrombin; T_B and F1 are degradative products of the thrombin enzyme.
Reprinted with permission from J. Rosing et al., *J. Biol. Chem.*, **255**, 274 (1980).

the migration toward the positive pole, polyacrylamide acts as a molecular sieve and the protein–micelle complexes are separated by size. As the proteins are denatured, artifacts caused by nonspheroid shapes of protein native structures will not be significant (Figure 2.48).

It should be realized that the detergents dissociate quaternary structure into its constituent subunits, and only the molecular weights of covalent subunits are determined by this method.

Separation of Proteins by Affinity Chromatography

Proteins have a high affinity to their substrates, prosthetic groups, membrane receptors, specific noncovalent inhibitors, and specific antibodies made against the protein. These high affinity compounds can be covalently attached to an insoluble resin and utilized to purify a protein in column chromatography. In a mixture of compounds eluted through the resin, the protein of interest will be selectively retarded.

General Approach to Protein Purification

A protein must be purified prior to a meaningful characterization of its chemical composition, structure, and function. As living cells contain thousands of genetically distinct proteins, the purification of a single protein from a mixture of cellular molecules may be difficult.

The first task in the purification of a protein is the development of a facile assay for the protein. A protein assay, whether it utilizes the rate of a substrate's transformation to a product, an antibody–antigen reaction, or a physiological response in an animal assay system, must in some way give a quantitative measurement of activity per unit weight for a particular sample containing the protein. This quantity is known as the sample's specific activity. The purpose of the purification is to increase a sample's specific activity to the value equal to the specific activity expected for the pure protein.

A typical protocol for purification of a soluble cellular protein may first involve the disruption of the cellular membrane, followed by a differential centrifugation in a density gradient to isolate the protein activity from subcellular particles and high molecular weight aggregates. A further purification step may utilize selective precipitation by addition of inorganic salts (salting out) or addition of miscible organic solvent to the solution containing the protein. Final purification will include a combination of techniques previously discussed, which include methods based on molecular charge (Section 2.3), molecular size, and affinity chromatography.

BIBLIOGRAPHY

Protein Structure

Anfinsen, C. B., and Scheraga, H. A. Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* 29:205, 1975.

Chou, P. Y., and Fasman, G. D. Empirical predication of protein conformation. *Annu. Rev. Biochem.* 47:251, 1978.

Dickerson, R. E., and Geis, I. *The structure and action of proteins*. New York: Harper & Row, 1969.

Neurath, H., and Hill, R. L. (eds.). *The proteins*, 3rd ed., vols. 1–4. New York: Academic, 1975–79.

Privalov, P. L. Stability of proteins. *Adv. Protein Chem.* 33:167, 1979.

Richards, F. M., and Richmond, T. Solvents, interfaces and protein structure, in *Molecular interactions and activity in proteins*, Ciba Foundation Symposium 60 (new series). Amsterdam: Excerpta Medica, 1978, p. 23.

Amino Acid Analysis and Sequencing

Hirs, C. H. W., and Timosheff, S. N. (eds.). *Methods Enzymol.* 47, 1977.

Motion In Folded Proteins

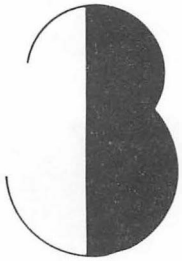
Huber, R. Conformational flexibility and its functional significance in some protein molecules. *Trends Biochem. Res.* 4:271, 1979.

McCammon, J. A., Gelin, B. R., and Karplus, M. Dynamics of folded proteins. *Nature.* 267:585, 1977.

Wüthrich, K., and Wagner, G. Internal motion in globular proteins. *Trends Biochem. Res.* 3:227, 1978.

Protein Spectroscopy

D'Albis, A., and Gratzer, W. B. Electronic spectra and optical activity of proteins, in A. T. Bull, J. R. Lagnodo, J. O. Thomas, and K. F. Tipton (eds.), *Companion to biochemistry*. London: Longmans, 1974.



Proteins II

RICHARD M. SCHULTZ

PHYSIOLOGICAL PROTEINS

3.1 OVERVIEW

In this chapter several examples of proteins with important biological functions are discussed. It is hoped that the particular examples chosen will serve as models from which the reader can gain general knowledge of protein structure and how these structural properties are related to important physiological functions of proteins. The ultimate objective is to understand how proteins function on an atomic level. Among the proteins chosen for discussion are some of the most investigated, such as hemoglobin, myoglobin, and the immunoglobulins, for which volumes of research papers have been written. However, even for these most investigated of proteins, the study of their architecture and how it relates to their function remains an extremely active area for research.

Section 3.2 discusses hemoglobin and myoglobin, which are examples of globular proteins. *Globular proteins* are spherically shaped, water-soluble proteins with a varied range of molecular weights and degrees of regular secondary structure. Hemoglobin and myoglobin are particularly interesting examples, and we have sufficient knowledge of these proteins to attempt a description of their physiological properties based on their molecular structures.

In Section 3.3 the important group of physiological proteins known as the *immunoglobulins* or *antibodies* are discussed. The immunoglobulins are found in high concentration in blood, comprising the major part of the γ_1 and γ_2 peaks in the electrophoresis pattern of plasma (see Figure 2.11).

The plasma, of course, contains many other proteins of physiological significance. The electrophoresis pattern of plasma represents hundreds of distinct proteins, some of which are present in rather small amounts. Table 3.1 lists only the major types of proteins in plasma for which functions are known. The *plasma lipoproteins,* discussed in Section 3.4, represent a distinctive part of the α and β electrophoresis peaks of the plasma proteins. These lipoproteins perform an essential role in the transport and metabolism of lipids in the human being. Their structure and properties may serve as models for other important groups of lipoproteins, such as those found in biological membranes.

In Section 3.5 the *glycoproteins* are generally described, in order to give the reader some knowledge of the structures of this important protein type.

In Section 3.6 examples from the class of *fibrous proteins* are described. These proteins differ from the globular type of protein in their relatively low water solubility, higher amount of secondary structure, elongated "rodlike" shape, high tensile strength, and, in

Table 3.1 Major Plasma Proteins Classified According to Function

Functions	*Protein*	*Normal Concentration (adults, mg/100 ml serum)*
TRANSPORT		
Cholesterol, Esterified fatty acids, Phospholipids	Plasma lipoproteins	α, 290–770; β, 190–600
Cortisol	Transcortin	~7
Vitamin B_{12}	Transcobalamin	—
Retinol	Retinol-binding protein	3–6
Hemin	Hemopexin	50–115
Hemoglobin	Haptoglobin	100–300
Iron	Transferrin	200–400
Copper	Ceruloplasmin	15–60
Aliphatic and aromatic anions	Albumin	3500–5500
Thyroxine	Thyroxine-binding globulin	1–2
	Albumin	3500–5500
	Prealbumin	10–40
DEFENSE		
Defense against Bacteria, Viruses, Foreign protein	Complement system	At least 13 components
	Immunoglobulin	At least 5 components
	Interferon	
Defense against blood loss	Clotting proteins	At least 10 components
	Prothrombin	5–10
	Plasminogen	10–30
	Fibrinogen	200–450 in plasma
Inhibitors against raised enzyme levels	α_1-Antitrypsin	200–400
	α_1-Antichymotrypsin	30–60
	α_2-Macroglobulin	150–420
	CI-Inactivator	15–35
	Antithrombin III	17–30
OTHER FUNCTIONS		
Fibrin lysis	Plasminogen	10–30
Muscle relaxant ?	Cholinesterase	0.5–1.5
Immunosuppressin	α_1-Fetoprotein	–
	α_1-Acid glycoprotein	55–140
Promotion of phagocytosis	C-reactive protein	<1.2
Antigenicity (HLA antigens)	α_2-Microglobulin	0.2
Source of protein during acute starvation	Albumin	3500–5500

Data taken from M. J. Geisow and A. H. Gordon, *Trends Biochem. Sci.* 3, 169 (1978) and F. W. Putnam, in *The Plasma Proteins*, Vol. I, 2nd ed., F. W. Putnam, ed., Academic Press, New York, 1975, Chapter I.

some cases, their elasticity. They provide a structural matrix for the organs and tissues of the mammalian organism.

3.2 HEMOGLOBIN AND MYOGLOBIN

The hemoglobins are globular proteins, in high concentration in red blood cells, that bind oxygen in the lungs and transport the oxygen in blood to the tissues and cells about the capillary beds of the vascular system. On returning to the lungs from the capillary beds, hemoglobins act to transport CO_2 and protons. In this section the structural and molecular aspects of the hemoglobin and myoglobin molecules are described. The physiological role of these proteins in gas transport is described in Chapter 23.

Forms of Human Hemoglobin

A hemoglobin molecule consists of four polypeptide chains of two different primary structures. In the common form of human adult hemoglobin, Hb A_1, two chains of one kind are designated the α chains, and the second two chains of the same kind are designated the β chains. The polypeptide chain composition of Hb A_1 is therefore $\alpha_2\beta_2$. The α-polypeptide chains contain 141 amino acids, and the β-polypeptide chains contain 146 amino acids.

While Hb A_1 is the major form of hemoglobin in the adult human, other forms of hemoglobin predominate in the blood of the human fetus. These early forms of human hemoglobin contain two of the same α chains found in Hb A_1, but their second kind of chain in the tetramer molecule differs in amino acid sequence from that of the β chain of adult Hb A_1 (Table 3.2). A minor form of adult hemoglobin, Hb A_2, comprises about 2% of normal adult hemoglobin and is composed of two α chains and two chains designated *delta* (δ) (Table 3.2).

Table 3.2 Chains of Human Hemoglobin

Primary Source	*Symbol*	*Chains*
Adult	Hb A_1	$\alpha_2\beta_2$
Adult	Hb A_2	$\alpha_2\delta_2$
Fetus	Hb F	$\alpha_2\gamma_2$
Embryonic (early fetus)	Hb Graves	$\alpha_2\varepsilon_2$

Myoglobin

Myoglobin (Mb) is an O_2-carrying protein that binds and releases O_2 with changes in the oxygen concentration in the cytoplasm of muscle cells. In contrast to hemoglobin, which has four polypeptide chains and four O_2 binding sites, myoglobin is composed of only a single polypeptide chain and a single O_2 binding site.

A comparison between myoglobin and hemoglobin is instructive in that myoglobin is a model for what occurs when a single protomer molecule acts alone without the interactions exhibited among the four O_2 binding sites in the more complex tetramer molecule of hemoglobin.

Heme Prosthetic Group

The four polypeptide chains in hemoglobin and the single polypeptide chain of myoglobin contain a heme prosthetic group. As defined in Chapter 2, a prosthetic group is a nonpolypeptide moiety that forms a part of a protein in its native functional state. A protein without its prosthetic group is designated an apoprotein.

The heme is a porphyrin molecule containing an iron atom in its center. The type of porphyrin found in hemoglobin and myoglobin (protoporphyrin IX) contains two propionic acid, two vinyl, and four methyl side-chain groups attached to the pyrrole rings of the porphyrin structure (Figure 3.1). The iron atom is in the ferrous (+2 charge) oxidation state in functional hemoglobin and myoglobin.

The ferrous atom in the heme can form five or six ligand bonds, depending on whether or not O_2 is bound to the protein. Of the five or six bonds, four of these bonds are to the pyrrole nitrogens of the porphyrin. Since all the pyrrole rings of the porphyrin lie in a com-

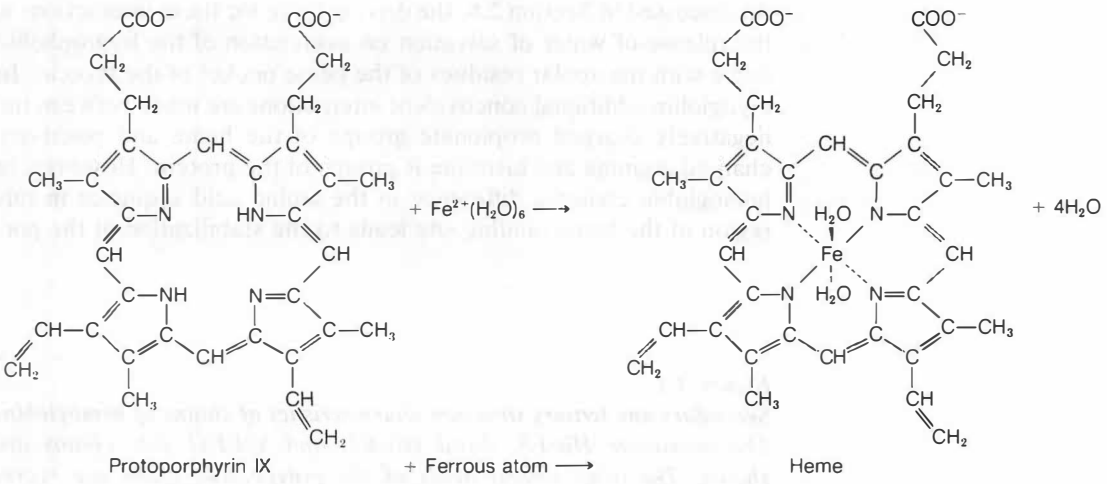


Figure 3.1
Structure of heme.

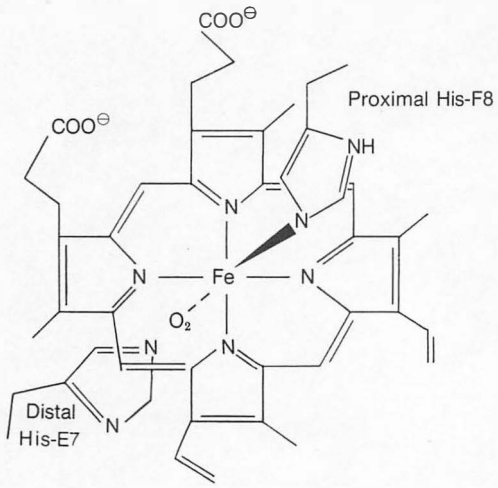


Figure 3.2
The ligand bonds to the ferrous atom in oxyhemoglobin.

mon plane, the four ligand bonds from the porphyrin to the iron atom at its center will also have a tendency to lie in the plane of the porphyrin ring. This is especially true for six-coordinate ferrous iron in the oxy form of hemoglobin. (In a later section it is described how the five-coordinate bonded ferrous atom of deoxyhemoglobin sits out of the plane of the porphyrin rings by about 0.6 Å.) The fifth and the potentially sixth ligand bonds to the ferrous atom of the heme are directed along an axis perpendicular to the plane of the porphyrin ring (Figure 3.2).

The fifth coordinate bond of the ferrous atom in each of the hemes is to a nitrogen of a histidine imidazole. This histidine is designated the *proximal histidine* in the hemoglobin and myoglobin structures (Figure 3.3).

In each of the polypeptide chains with O_2 bound, the O_2 molecule forms a sixth coordinate bond to the ferrous atom. In this bonded position the O_2 is placed between the ferrous atom to which it is liganded and a second histidine imidazole, designated the *distal histidine,* in hemoglobin and myoglobin structures. In deoxyhemoglobin, the sixth coordinate position (O_2 binding position) of the ferrous atom is unoccupied.

The porphyrin part of the heme is positioned within a hydrophobic pocket formed in each of the polypeptide chains. In the heme pocket x-ray crystal diffraction structures show that approximately 80 interactions are provided by approximately 15 amino acids to the heme. Most of these noncovalent interactions are between apolar side chains of amino acids and the nonpolar regions of the porphyrin. As discussed in Section 2.6, the driving force for these interactions is the release of water of solvation on association of the hydrophobic heme with the apolar residues of the heme pocket of the protein. In myoglobin additional noncovalent interactions are made between the negatively charged propionate groups of the heme and positively charged arginine and histidine R groups of the protein. However, in hemoglobin chains a difference in the amino acid sequence in this region of the heme binding site leads to the stabilization of the por-

Figure 3.3
Secondary and tertiary structure characteristics of chains of hemoglobin. *The proximate His-F8, distal His-E7, and Val-E11 side chains are shown. The other amino acids of the polypeptide chain are represented by their α-carbon positions only; the letters M, V, and P refer to the methyl, vinyl, and propionate side chains of the heme.*
Reprinted with permission from M. Perutz, *Brit. Med. Bull.*, 32, 195 (1976).

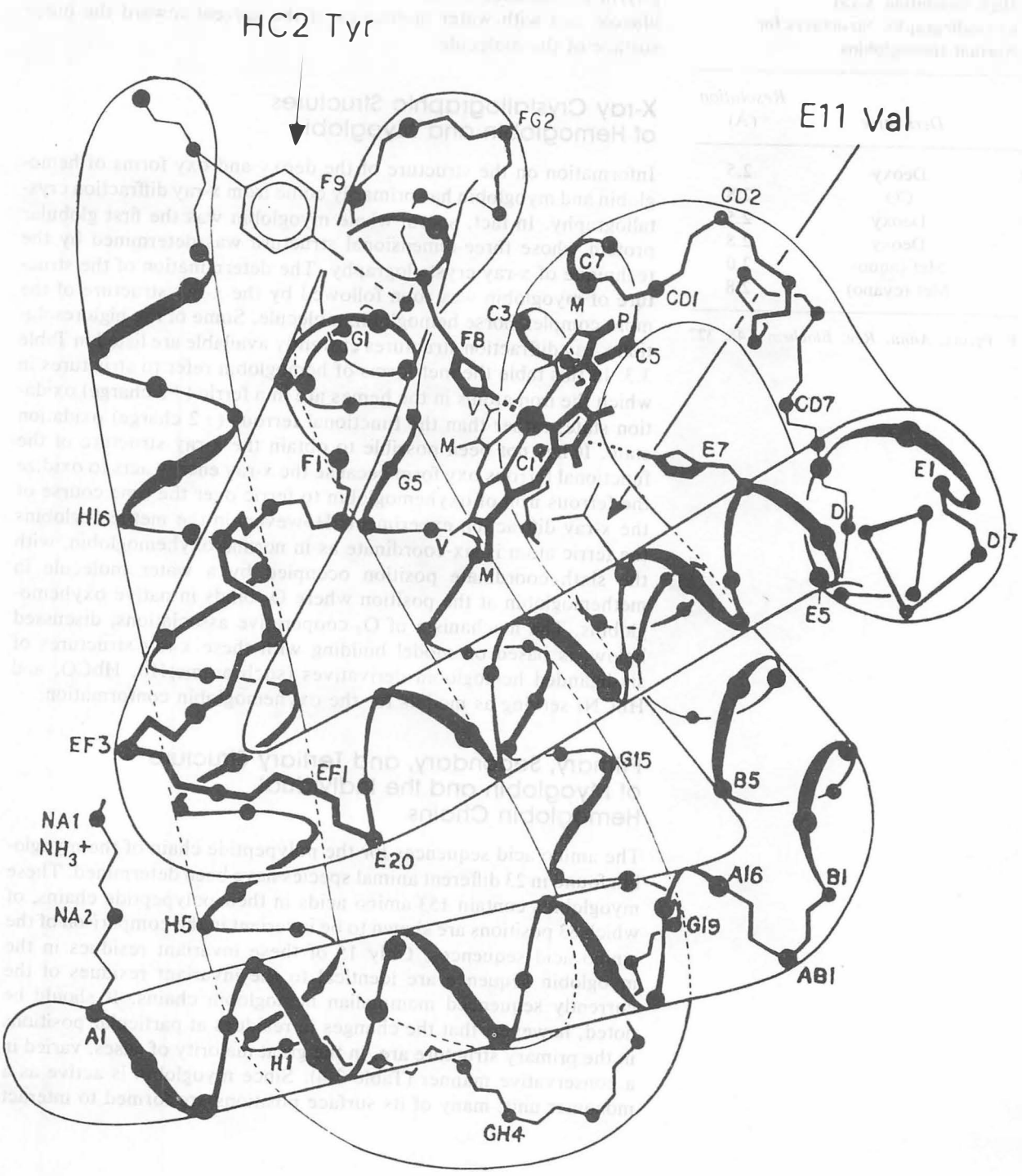

HC2 Tyr
E11 Val
FG2
F9
CD2
C7
CD1
C3
M
P
G1
F8
C5
V
CD7
M
C1
E7
E1
F1
G5
H16
D1
D7
V
M
E5
EF3
G15
EF1
B5
NA1
NH3+
E20
A16
B1
NA2
H5
G19
AB1
A1
H1
GH4

Table 3.3 High Resolution X-ray Crystallographic Structures for Normal Hemoglobins

Species	*Derivative*	*Resolution (Å)*
Human A	Deoxy	2.5
	CO	2.8
Human F	Deoxy	2.5
Horse	Deoxy	2.8
	Met (aquo)	2.0
	Met (cyano)	2.8

From M. F. Perutz, *Annu. Rev. Biochem.*, **48**, 327 (1979).

phyrin propionates by interaction with an uncharged histidine imidazole and with water molecules of the solvent toward the outer surface of the molecule.

X-ray Crystallographic Structures of Hemoglobin and Myoglobin

Information on the structure of the deoxy and oxy forms of hemoglobin and myoglobin has primarily come from x-ray diffraction crystallography. In fact, sperm whale myoglobin was the first globular protein whose three-dimensional structure was determined by the technique of x-ray crystallography. The determination of the structure of myoglobin was soon followed by the x-ray structure of the more complex horse hemoglobin molecule. Some of the high resolution x-ray diffraction structures currently available are listed in Table 3.3. In this table the met forms of hemoglobin refer to structures in which the iron atoms in the hemes are in a ferric (+3 charge) oxidation state, rather than the functional ferrous (+2 charge) oxidation state. It has not been possible to obtain the x-ray structure of the functional ferrous oxy form because the x-ray energy acts to oxidize the ferrous iron of oxyhemoglobin to ferric over the time course of the x-ray diffraction experiment. However, in the methemoglobins the ferric atom is six-coordinate as in normal oxyhemoglobin, with the sixth coordinate position occupied by a water molecule in methemoglobin at the position where O_2 binds in native oxyhemoglobins. The mechanism of O_2 cooperative associations, discussed below, is based on model building with these x-ray structures of six-liganded hemoglobin derivatives (such as metHb, HbCO, and HbCN) serving as models for the oxyhemoglobin conformation.

Primary, Secondary, and Tertiary Structure of Myoglobin and the Individual Hemoglobin Chains

The amino acid sequences for the polypeptide chain of the myoglobin found in 23 different animal species have been determined. These myoglobins contain 153 amino acids in their polypeptide chains, of which 83 positions are shown to be invariant in the comparison of the amino acid sequences. Only 15 of these invariant residues in the myoglobin sequence are identical to the invariant residues of the currently sequenced mammalian hemoglobin chains. It should be noted, however, that the changes in residues at particular positions in the primary structure are, in the great majority of cases, varied in a conservative manner (Table 3.4). Since myoglobin is active as a monomer unit, many of its surface positions are formed to interact

Table 3.4 Amino Acid Sequences of Human Hemoglobin Chains and of Sperm Whale Myoglobin[a]

		NA 1	2	3	A 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	A 16	AB 1	B 1	2	3	4	5	6
MYOGLOBIN		val	.....	leu	ser	glu	gly	glu	trp	gln	leu	val	leu	his	val	trp	ala	lys	val	glu	ala	asp	val	ala	gly	his	gly
HEMOGLOBIN	Horse α	val	.....	leu	ser	ala	ala	asp	lys	thr	asn	val	lys	ala	ala	trp	ser	lys	val	gly	gly	his	ala	gly	glu	tyr	gly
	β	val	gln	leu	ser	gly	glu	glu	lys	ala	ala	val	leu	ala	leu	trp	asp	lys	val	asn	.....	.....	glu	glu	glu	val	gly
	Human α	val	.....	leu	ser	pro	ala	asp	lys	thr	asn	val	lys	ala	ala	trp	gly	lys	val	gly	ala	his	ala	gly	glu	tyr	gly
	β	val	his	leu	thr	pro	glu	glu	lys	ser	ala	val	thr	ala	leu	trp	gly	lys	val	asn	.....	.....	val	asp	glu	val	gly
	γ	gly	his	phe	thr	glu	glu	asp	lys	ala	thr	ilu	thr	ser	leu	trp	gly	lys	val	asn	.....	.....	val	glu	asp	ala	gly
	δ	val	his	leu	thr	pro	glu	glu	lys	thr	ala	val	asn	ala	leu	trp	gly	lys	val	asn	.....	.....	val	asp	ala	val	gly

		7	8	9	10	11	12	13	14	15	16	C 1	2	3	4	5	6	7	CD 1	2	3	4	5	6	7	8	D 1
MYOGLOBIN		gln	asp	ilu	leu	ilu	arg	leu	phe	lys	ser	his	pro	glu	thr	leu	glu	lys	phe	asp	arg	phe	lys	his	leu	lys	thr
HEMOGLOBIN	Horse α	ala	glu	ala	leu	glu	arg	met	phe	leu	gly	phe	pro	thr	thr	lys	thr	tyr	phe	pro	his	phe	.....	asp	leu	ser	his
	β	gly	glu	ala	leu	gly	arg	leu	leu	val	val	tyr	pro	trp	thr	gln	arg	phe	phe	asp	ser	phe	gly	asp	leu	ser	gly
	Human α	ala	glu	ala	leu	glu	arg	met	phe	leu	ser	phe	pro	thr	thr	lys	thr	tyr	phe	pro	his	phe	.....	asp	leu	ser	his
	β	gly	glu	ala	leu	gly	arg	leu	leu	val	val	tyr	pro	trp	thr	gln	arg	phe	phe	glu	ser	phe	gly	asp	leu	ser	thr
	γ	gly	glu	thr	leu	gly	arg	leu	leu	val	val	tyr	pro	trp	thr	gln	arg	phe	phe	asp	ser	phe	gly	asn	leu	ser	ser
	δ	gly	glu	ala	leu	gly	arg	leu	leu	val	val	tyr	pro	trp	thr	gln	arg	phe	phe	glu	ser	phe	gly	asp	leu	ser	ser

		2	3	4	5	6	7	E 1	2	3	4	5	6	7	8	9	10	11	12	13	14	E 15	16	17	18	19	20
MYOGLOBIN		glu	ala	glu	met	lys	ala	ser	glu	asp	leu	lys	lys	his	gly	val	thr	val	leu	thr	ala	leu	gly	ala	ilu	leu	lys
HEMOGLOBIN	Horse α	.....	.....	.....	.....	.....	gly	ser	ala	gln	val	lys	ala	his	gly	lys	lys	val	ala	asp	gly	leu	thr	leu	ala	val	gly
	β	pro	asp	ala	val	met	gly	asn	pro	lys	val	lys	ala	his	gly	lys	lys	val	leu	his	ser	phe	gly	glu	gly	val	his
	Human α	.....	.....	.....	.....	.....	gly	ser	ala	gln	val	lys	gly	his	gly	lys	lys	val	ala	asp	ala	leu	thr	asn	ala	val	ala
	β	pro	asp	ala	val	met	gly	asn	pro	lys	val	lys	ala	his	gly	lys	lys	val	leu	gly	ala	phe	ser	asp	gly	leu	ala
	γ	ala	ser	ala	ilu	met	gly	asn	pro	lys	val	lys	ala	his	gly	lys	lys	val	leu	thr	ser	leu	gly	asp	ala	ilu	lys
	δ	pro	asp	ala	val	met	gly	asn	pro	lys	val	lys	ala	his	gly	lys	lys	val	leu	gly	ala	phe	ser	asp	gly	leu	ala

		EF 1	2	3	4	5	6	7	8	F 1	2	3	4	F 5	6	7	8	9	FG 1	2	3	4	5	G 1	2	3	4
MYOGLOBIN		lys	lys	gly	his	his	glu	ala	glu	leu	lys	pro	leu	ala	gln	ser	his	ala	thr	lys	his	lys	ilu	pro	ilu	lys	tyr
HEMOGLOBIN	Horse α	his	leu	asp	asp	leu	pro	gly	ala	leu	ser	asp	leu	ser	asn	leu	his	ala	his	lys	leu	arg	val	asp	pro	val	asn
	β	his	leu	asp	asn	leu	lys	gly	thr	phe	ala	ala	leu	ser	glu	leu	his	cys	asp	lys	leu	his	val	asp	pro	glu	asn
	Human α	his	val	asp	asp	met	pro	asn	ala	leu	ser	ala	leu	ser	asp	leu	his	ala	his	lys	leu	arg	val	asp	pro	val	asn
	β	his	leu	asp	asn	leu	lys	gly	thr	phe	ala	thr	leu	ser	glu	leu	his	cys	asp	lys	leu	his	val	asp	pro	glu	asn
	γ	his	leu	asp	asp	leu	lys	gly	thr	phe	ala	gln	leu	ser	glu	leu	his	cys	asp	lys	leu	his	val	asp	pro	glu	asn
	δ	his	leu	asp	asn	leu	lys	gly	thr	phe	ser	gln	leu	ser	glu	leu	his	cys	asp	lys	leu	his	val	asp	pro	glu	asn

		5	6	7	8	G 9	10	11	12	13	14	15	16	17	18	19	GH 1	2	3	4	5	6	H 1	2	H 3	4	5
MYOGLOBIN		leu	glu	phe	ilu	ser	glu	ala	ilu	ilu	his	val	leu	his	ser	arg	his	pro	gly	asn	phe	gly	ala	asp	ala	gln	gly
HEMOGLOBIN	Horse α	phe	lys	leu	leu	ser	his	cys	leu	leu	ser	thr	leu	ala	val	his	leu	pro	asn	asp	phe	thr	pro	ala	val	his	ala
	β	phe	arg	leu	leu	gly	asn	val	leu	ala	leu	val	val	ala	arg	his	phe	gly	lys	asp	phe	thr	pro	glu	leu	gln	ala
	Human α	phe	lys	leu	leu	ser	his	cys	leu	leu	val	thr	leu	ala	ala	his	leu	pro	ala	glu	phe	thr	pro	ala	val	his	ala
	β	phe	arg	leu	leu	gly	asn	val	leu	val	cys	val	leu	ala	his	his	phe	gly	lys	glu	phe	thr	pro	pro	val	gln	ala
	γ	phe	lys	leu	leu	gly	asn	val	leu	val	thr	val	leu	ala	ilu	his	phe	gly	lys	glu	phe	thr	pro	glu	val	gln	ala
	δ	phe	arg	leu	leu	gly	asn	val	leu	val	cys	val	leu	ala	arg	asn	phe	gly	lys	glu	phe	thr	pro	gln	met	gln	ala

		6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	H 21	22	23	2	HC 1	2	3	4	5
MYOGLOBIN		ala	met	asn	lys	ala	leu	glu	leu	phe	arg	lys	asp	ilu	ala	ala	lys	tyr	lys	glu	leu	gly	tyr	gln	gly
HEMOGLOBON	Horse α	ser	leu	asp	lys	phe	leu	ser	ser	val	ser	thr	val	leu	thr	ser	lys	tyr	arg						
	β	ser	tyr	gln	lys	val	val	ala	gly	val	ala	asn	ala	leu	ala	his	lys	tyr	his						
	Human α	ser	leu	asp	lys	phe	leu	ala	ser	val	ser	thr	val	leu	thr	ser	lys	tyr	arg						
	β	ala	tyr	gln	lys	val	val	ala	gly	val	ala	asn	ala	leu	ala	his	lys	tyr	his						
	γ	ser	trp	gln	lys	met	val	thr	gly	val	ala	ser	ala	leu	ser	ser	arg	tyr	his						
	δ	ala	tyr	gln	lys	val	val	ala	gly	val	ala	asn	ala	leu	ala	his	lys	tyr	his						

Based on diagram in R. E. Dickerson and I. Geis, *The Structure and Function of Proteins*, Harper & Row, New York, 1969, p. 52.

[a] Residues that are identical are enclosed in box. A, B, C . . . designate different helices of tertiary structure (see text).

with water and prevent another molecule of myoglobin from associating. The surface residues of the individual chains in hemoglobin are designed to provide hydrogen bonds and nonpolar contacts with other subunits in the hemoglobin quaternary structure. The proximal and distal histidines are, of course, preserved in the sequences of all the polypeptide chains. Other invariant residues are in the hydrophobic heme pocket and form essential nonpolar contacts with the heme that stabilizes the heme–protein complex. In addition, prolines, which act to break some of the helical sections to allow the chain to fold back upon itself, are predominantly retained in most of the chains.

While there is a surprising variability in amino acid sequence among the different polypeptide chains, to a first approximation the secondary and tertiary structures of each of the polypeptide chains of hemoglobin and myoglobin appear to be almost identical (Figure 3.4). The significant differences in the physiological properties be-

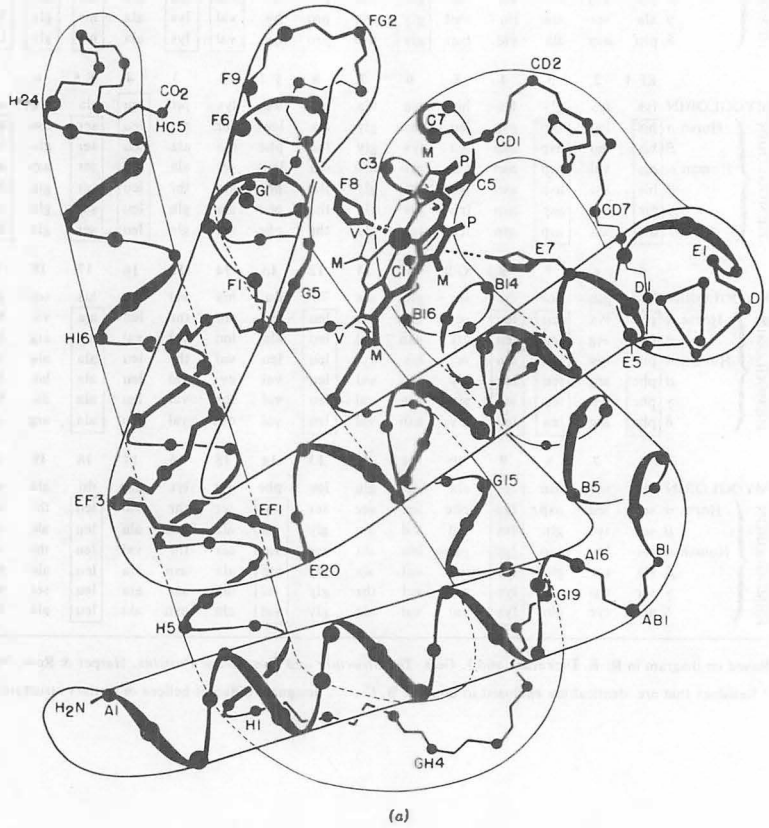
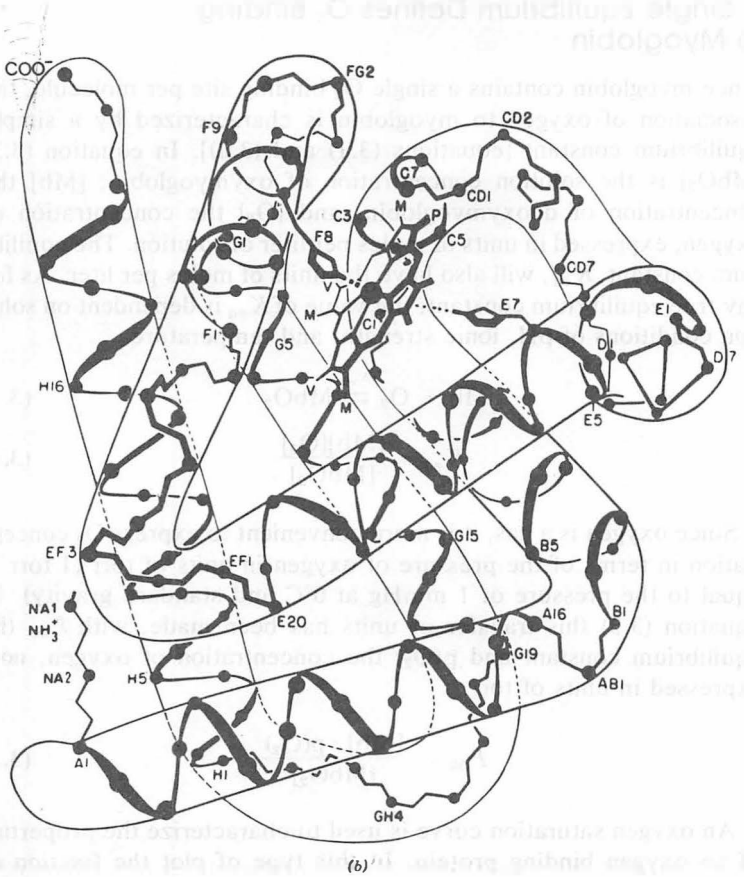


Figure 3.4
Comparison of the conformation of myoglobin (a) and β chain of Hb A_1 (b).
The overall structures are very similar, except at the NH_2-terminal and COOH-terminal ends.

Reprinted with permission from A. Fersht, *Enzyme Structure and Mechanism*, W. H. Freeman, San Francisco, Calif., 1977, pp. 12 and 13.

tween the α, β, γ, and δ chains in the hemoglobins and the single polypeptide chains of myoglobin are, therefore, due to rather small specific changes in the overall tertiary structure. The similarity in tertiary structure, resulting from widely varied amino acid sequences, show that the same tertiary structure for a protein can be arrived at in many different ways.

The x-ray crystallographic structure shows that each of the polypeptide chains are composed of multiple α-helical structures that are broken by turns of the polypeptide chain, allowing the protein to fold into a spheroid shape (Figure 3.4). Approximately 70% of the residues in the protein participate in the α-helical secondary structure, which generates seven helical segments in the α chain and eight helical segments in the β chain. These latter eight helical regions are commonly lettered A through H, starting from the A helix at the NH_2-terminal end, and the interhelical regions designated as AB, BC, CD, . . . , GH, respectively. The nonhelical region that lies



(b)

between the NH_2-terminal end and the A helix is designated the NA region; and the region between the COOH-terminal end and the H helix is designated the HC region (Figure 3.3). This identification system allows discussion of particular residues that have similar functional and structural roles in each of the different hemoglobin and myoglobin chains, even with deletions and additions of amino acids to regions of the primary structure, which must be made because the chains are of different lengths. The proximal histidine in all chains is residue F8 (eighth residue of the F helix).

In a comparison of the α and β chains of Hb A_1, the α chain differs from the β chain by the deletion of one amino acid residue from the NA segment, one residue in the CD region, and the deletion of five residues in the D helix. The α chain contains two "extra" residues at the AB region that are absent in the β chain. There are 80 amino acid differences throughout the two sequences (Table 3.4).

A Single Equilibrium Defines O_2 Binding to Myoglobin

Since myoglobin contains a single O_2 binding site per molecule, the association of oxygen to myoglobin is characterized by a simple equilibrium constant [equations (3.1) and (3.2)]. In equation (3.2) [MbO_2] is the solution concentration of oxymyoglobin, [Mb] the concentration of deoxymyoglobin, and [O_2] the concentration of oxygen, expressed in units of moles per liter of solution. The equilibrium constant, K_{eq}, will also have the units of moles per liter. As for any true equilibrium constant, the value of K_{eq} is dependent on solution conditions of pH, ionic strength, and temperature.

$$Mb + O_2 \overset{K_{eq}}{\rightleftharpoons} MbO_2 \tag{3.1}$$

$$K_{eq} = \frac{[Mb][O_2]}{[MbO_2]} \tag{3.2}$$

Since oxygen is a gas, it is more convenient to express O_2 concentration in terms of the pressure of oxygen in units of torr (1 torr is equal to the pressure of 1 mmHg at 0°C and standard gravity). In equation (3.3) this transfer of units has been made, with P_{50} the equilibrium constant and $p(O_2)$ the concentration of oxygen, now expressed in units of torr.

$$P_{50} = \frac{[Mb] \cdot p(O_2)}{[MbO_2]} \tag{3.3}$$

An oxygen saturation curve is used to characterize the properties of an oxygen binding protein. In this type of plot the fraction of

oxygen binding sites in solution that contain oxygen [Y, equation (3.4)] is plotted on the ordinate vs the p(O_2) (oxygen concentration) on the abscissa. The Y value is simply defined for myoglobin by equation (3.5). Substitution into equation (3.5) the value of [MbO_2] obtained from equation (3.3) and then dividing through by [Mb], results in equation (3.6), which shows the dependency of Y on the value of the equilibrium constant, P_{50}, and the oxygen concentration. It is seen from equations (3.3) and (3.6) that the value of P_{50} is equal to the oxygen concentration, p(O_2), when $Y = 0.5$ (50% of the available sites occupied). Hence the designation of the equilibrium constant by the subscript 50.

$$Y = \frac{\text{number of binding sites occupied}}{\text{total number of binding sites in solution}} \tag{3.4}$$

$$Y = \frac{[MbO_2]}{[Mb] + [MbO_2]} \tag{3.5}$$

$$Y = \frac{p(O_2)}{P_{50} + p(O_2)} \tag{3.6}$$

A plot of equation (3.6) of Y vs p(O_2) generates an oxygen saturation curve in the form of a rectangular hyperbola (Figure 3.5).

A simple algebraic manipulation of equation (3.6) leads to equation (3.7). Taking the logarithm of both sides of equation (3.7) results in equation (3.8), which is known as the Hill equation. A plot of log ($Y/1 - Y$) vs log p(O_2), according to equation (3.8), yields a straight line with a slope equal to 1 for myoglobin (Figure 3.6). This is called the Hill plot, and the slope (n_h) is referred to as the Hill coefficient [see equation (3.9)].

$$\frac{Y}{1 - Y} = \frac{p(O_2)}{P_{50}} \tag{3.7}$$

$$\log \frac{Y}{1 - Y} = \log p(O_2) - \log P_{50} \tag{3.8}$$

Binding of O_2 to Hemoglobin Is More Complex

Whereas myoglobin has a single O_2 binding site per molecule, hemoglobins contain a quaternary structure of four polypeptide chains, each with a heme binding site for O_2. The binding of the 4 O_2 in hemoglobin is found to be positively *cooperative*, so that the binding of the first O_2 to deoxyhemoglobin facilitates the binding of O_2 to the other subunits in the molecule. Conversely, the dissociation of the first O_2 from fully oxygenated hemoglobin [$Hb(O_2)_4$] will make

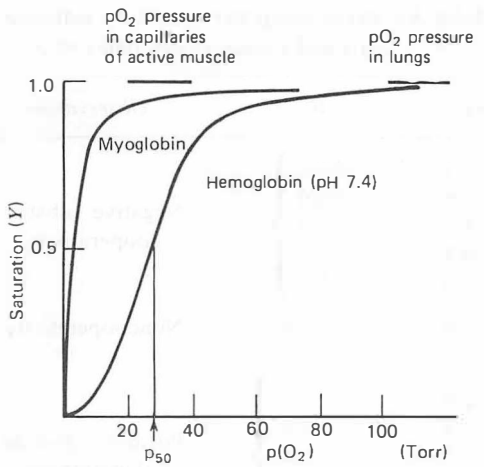


Figure 3.5
Oxygen binding curves for myoglobin and hemoglobin.

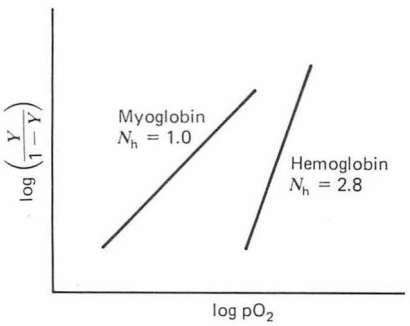


Figure 3.6
Hill plots for myoglobin and hemoglobin Hb A_1.

Table 3.5 Relationship Between Hill Coefficient (n_h) and Cooperativity Index (R_x)

n_h	R_x	*Observation*
0.5	6560	Negative substrate cooperativity
0.6	1520	
0.7	533	
0.8	243	
0.9	132	
1.0	81.0	Noncooperativity
1.5	18.7	Positive substrate cooperativity
2.0	9.0	
2.8	4.8	
3.5	3.5	
6.0	2.1	
10.0	1.6	
20.0	1.3	

Based on Table 7.1 in A. Cornish-Bowden, *Principals of Enzyme Kinetics*, Butterworths Scientific Publishers, London and Boston, 1976.

easier the dissociation of O_2 from the other subunits of the tetramer molecule.

Based on cooperativity in oxygen association and dissociation, the oxygen saturation curve for hemoglobin differs from that previously derived for myoglobin. A plot of Y vs $p(O_2)$ for hemoglobin follows a *sigmoid* line, indicating cooperativity in oxygen association (Figure 3.5). A plot of the Hill equation [equation (3.9)] gives a value of the slope (n_h) equal to 2.8.

$$\log \frac{Y}{1 - Y} = n_h \log p(O_2) - \text{Constant} \qquad (3.9)$$

The meaning of the Hill coefficient to cooperative O_2 association can be quantitatively evaluated in Table 3.5. A parameter known as the cooperativity index, R_x, is calculated, which shows the ratio of $p(O_2)$ concentrations required to change Y from a value of $Y = 0.1$ (10% of sites filled) to a value of $Y = 0.9$ (90% of sites filled) for designated Hill coefficient values found experimentally. In the case of myoglobin, $n_h = 1$, and an 81-fold change in oxygen concentration is required to change from $Y = 0.1$ to $Y = 0.9$. In hemoglobin, where positive cooperativity is observed, $n_h = 2.8$, and *only a 4.8-fold change in oxygen concentration* is required to change the fractional saturation from 0.1 to 0.9.

Molecular Mechanism of O_2 Cooperativity

The x-ray diffraction data on deoxyhemoglobin shows that the ferrous atoms sit out of the plane of their porphyrins by about 0.6 Å. This is thought to be due to two factors. The electronic configuration of five-coordinated ferrous atom in deoxyhemoglobin has a slightly larger radius (2.06 Å) than the distance from the center of the porphyrin to each of the pyrrole nitrogens (2.01 Å). Accordingly, the iron can be placed into the center of the porphyrin only with some distortion of the most stable porphyrin conformation. Probably a more important consideration is that if the iron atom sits in the plane of the porphyrin, the proximal His-F8 imidazole will interact unfavorably with atoms of the porphyrin. The strength of this unfavorable steric interaction is, in part, due to conformational constraints on the His-F8 and the porphyrin in the deoxyhemoglobin conformation that energetically forces the approach of the His-F8 toward the porphyrin to a particular path. These constraints will become less significant in the oxy conformation of hemoglobin. These repulsive interactions, however, are primarily responsible for the ferrous atom being approximately 0.6 Å out of the plane of the porphyrin (Figure 3.7).

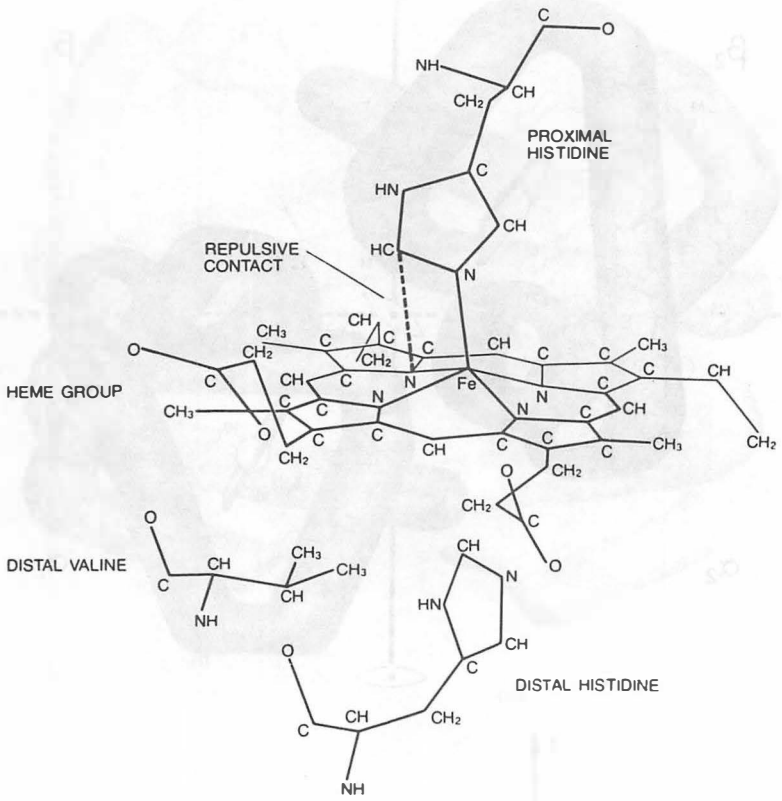


Figure 3.7
Steric hindrance between proximal histidine and porphyrin in deoxyhemoglobin.
From M. Perutz, *Sci. Am.*, 239, 92 (1978) by Scientific American, Inc. All rights reserved.

On the binding of O_2 to the first heme, the ferrous atom acquires its sixth coordinate bond. Corresponding with this change in coordination, the iron atom will overcome the repulsive interactions between the His-F8 and porphyrin and move into the plane of the porphyrin ring. This is the most thermodynamically stable position for the now six-bonded iron atom; one axial ligand is on either side of the plane of the porphyrin ring, and the steric repulsion of one of the axial ligands with the porphyrin is balanced by the repulsion of the second axial ligand on the opposite side when the ferrous atom is in the center. If the iron atom is displaced from the center, the steric interactions of the two axial ligands with the porphyrin in the deoxy conformation are unbalanced, and the stability of the unbalanced structure will be lower than that of the equidistant configuration. Also, the radius of the iron atom with six ligands is reduced to 1.94 Å, which can just fit into the center of the porphyrin without distortion of the porphyrin conformation.

As the steric repulsion between the porphyrin and the His-F8 must

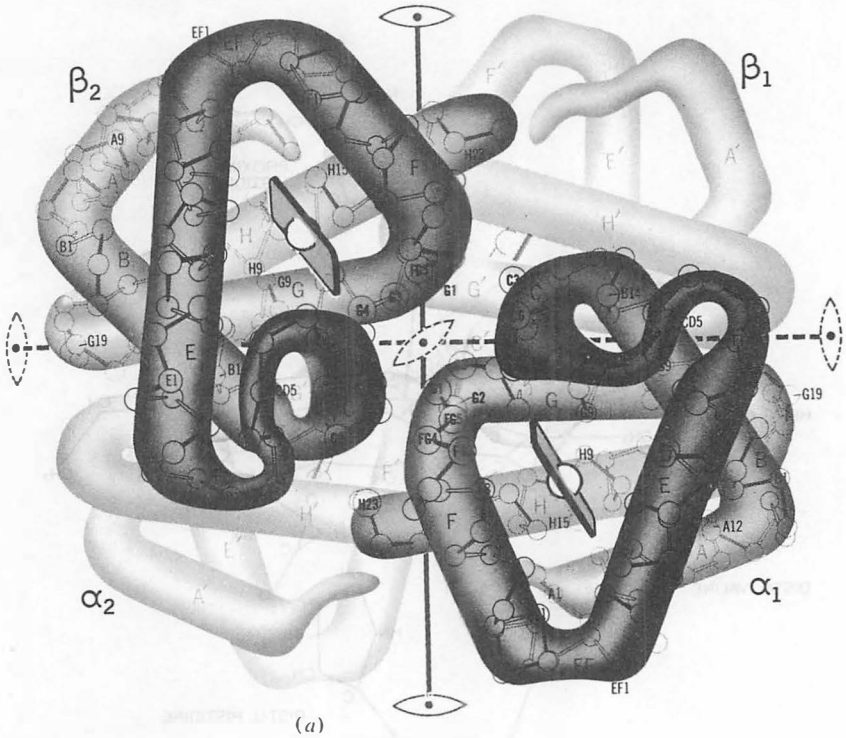


Figure 3.8

(a) Quaternary structure of hemoglobin showing the $\alpha_1\beta_2$ interface contacts between FG corners and C helix. (b) Cylinder representation of α_1 and β_2 subunits in hemoglobin molecule showing α_1 and β_2 interface contacts between FG corner and C helix.

(*a*) Figure reprinted with permission from R. E. Dickerson and I. Geis, *The Structure and Action of Proteins*, W. A. Benjamin, Inc., Menlo Park, 1969, p. 56. (*b*) Figure reprinted with permission from J. Baldwin and C. Chothia, *J. Mol. Biol.* **129**, 175 (1979). Copyright by Academic Press Inc. (London) Ltd.

be overcome on O_2 association, a significant energy is required for O_2 association, and the binding of the first O_2 to hemoglobin is characterized by a relatively low affinity constant. When an O_2 association does occur to one of the ferrous hemes, the change in the iron atom position from 0.6 Å above the plane of the porphyrin into the center of the porphyrin triggers a conformational change in the whole hemoglobin molecule. The change in conformation results in the greater facility (lower energy) for O_2 to combine with other heme sites after the first O_2 has bound. This conformation change is thought to occur as described below, based on information supplied by the x-ray-determined structures of the proteins listed in Table 3.3.

The conformation of the deoxyhemoglobin appears to be stabilized by interactions of the quaternary structure at the $\alpha_1\beta_2$ and $\alpha_2\beta_1$ subunit interfaces where the FG corner of one subunit interacts with the C helix of the other subunit (Figure 3.8). In addition, ionic interactions, especially with ionized groups at the COOH-terminal regions of the β chains may stabilize the deoxy conformation of the protein (Figure 3.9). These intersubunit interactions are destabilized on the binding of O_2 to one of the heme subunits of a hemoglobin molecule as follows. The movement of the His-F8 toward the porphyrin, correlated with the movement of the ferrous atom into the plane of the porphyrin on binding O_2, pulls the F helix with it. The movement of the F helix, in turn, moves the FG corner of the subunit, destabilizing its interaction with the C helix of the adjacent subunit at the $\alpha_1\beta_2$ or $\alpha_2\beta_1$ subunit interface.

The FG to C intersubunit contacts are thought to act as a "switch," because they apparently can exist in two different arrangements with different modes of contact between FG and C. On the binding of O_2 to deoxyhemoglobin, the movement of the FG corner in the subunit to which the O_2 is bound forces the disruption of favorable hydrophobic interactions that stabilize the deoxy conformation at the FG to C interface, and the intersubunit contacts

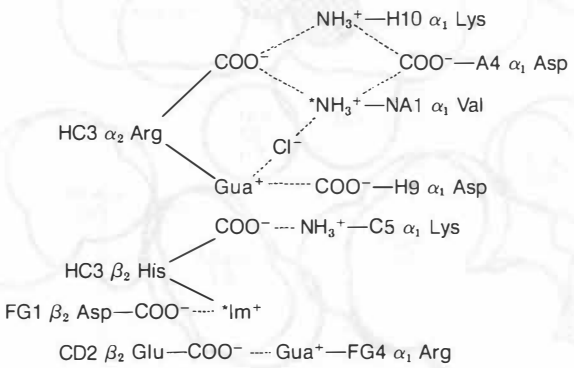


Figure 3.9
Salt bridges between subunits in deoxyhemoglobin.
Im^+ is imidazole; Gua^+ is guanidine; starred residues account for approximately 60% of the alkaline Bohr effect.
Redrawn from M. Perutz, *Brit. Med. Bull.*, **32**, 195 (1976).

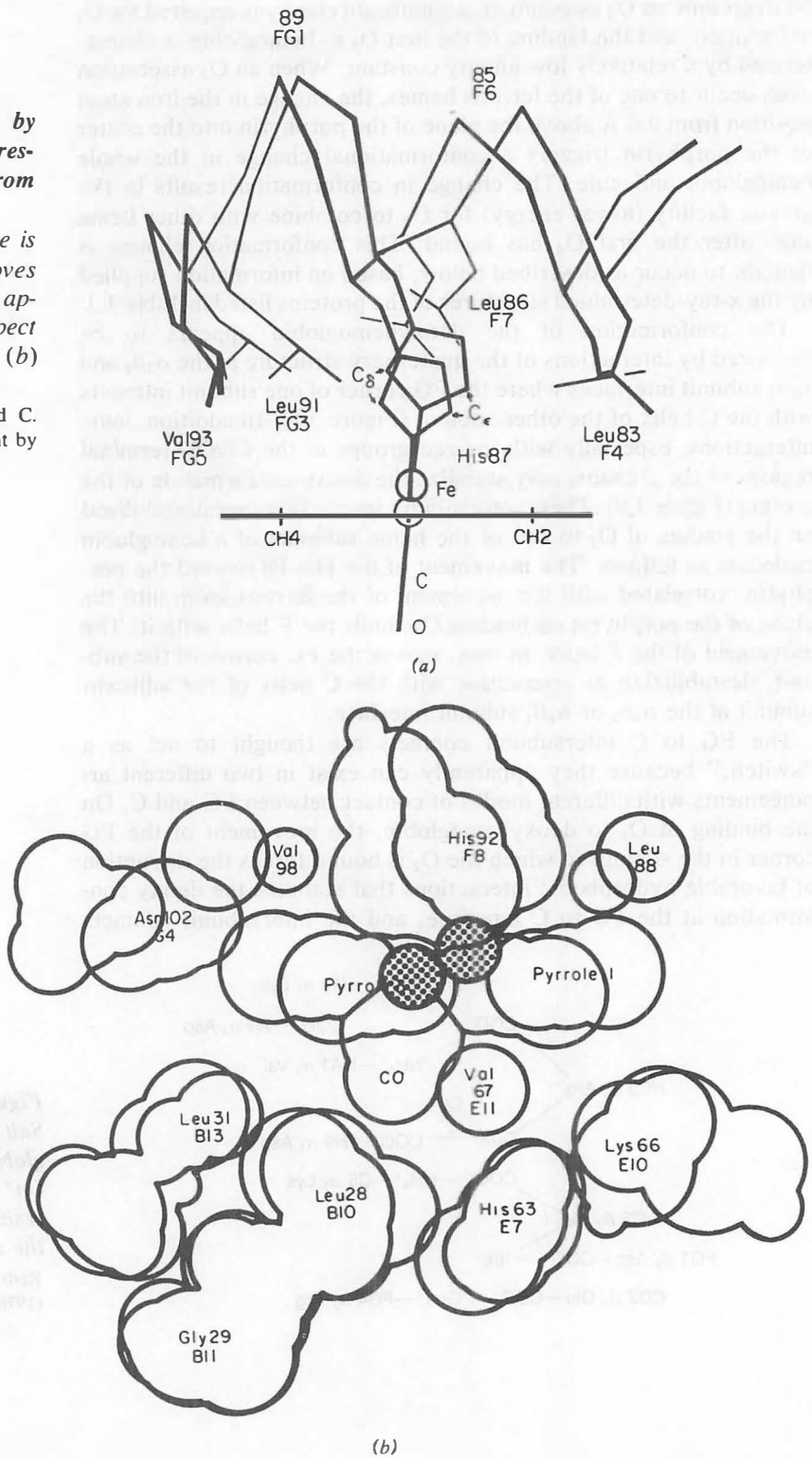


Figure 3.10

A stick and space-filling diagram drawn by computer graphics showing movements of residues in heme environment on transition from deoxyhemoglobin to carboxyhemoglobin.

(a) Movements in α subunit shown; heme is perpendicular to the figure. F helix moves toward heme and tilts so that His-F8 approach is more symmetrical with respect to heme on shift to oxy conformation. (b) Similar movement in β subunit shown.

switch to their alternative position. The switch between the two positions involves a relative movement of FG and C in opposite subunits of about 6 Å. In the second position of the "switch," the tertiary conformation of the subunits participating in the FG to C intersubunit contact are less constrained and can change to a new tertiary conformation (oxy conformation). This second conformation will allow the His-F8 to approach their porphyrins, on O_2 association, with a less significant steric repulsion than in the deoxy conformation of the hemoglobin subunits (Figure 3.10). Thus the O_2 molecule can bind to the heme in the less constrained oxy conformation more easily than to the original subunit conformation held by the intersubunit FG to C interaction in the deoxy conformation.

In addition, the Val-E11 in the deoxy conformation of the β subunits is at the entrance to the O_2 binding site, where it sterically impedes O_2 association to the heme. In the oxy conformation the heme in the β subunits appears to move approximately 1.5 Å further into the heme binding site of the protein, changing the geometric relationship of the O_2 binding site in the heme to the Val-E11 side chain, so that the Val-E11 no longer sterically interferes with O_2 binding. This appears to be an important additional factor that results in a higher affinity of O_2 to the oxy conformation of the β chain than to the deoxy conformation.

The deoxy conformation of hemoglobin is referred to as the "tight" or T conformational state. The oxyhemoglobin conformational form is referred to as the "relaxed" or R conformational state. On the binding of O_2 to the initial heme subunit of the tetramer molecule, the molecule is transformed from the T to the R conformational state. The affinity constant of O_2 is greater for the R state than the T state by a factor of 150 to 300, depending on the solution conditions.

Molecular Mechanism of the Bohr Effect

The equilibrium expression for oxygen association to hemoglobin includes a term that indicates the participation of hydrogen ion in the equilibrium.

$$Hb + 4O_2 \rightleftharpoons Hb(O_2)_4 + xH^+ \qquad (3.10)$$

Equation (3.10) shows that the R form is more acidic, and protons dissociate when the hemoglobin is transformed to the R form. The number of equivalents of protons that dissociate per mole of deoxyhemoglobin depends on the pH of the solution; 2.8 equiv protons/mol Hb are produced at pH 7.4. The production of protons at alkaline pH (pH > 6), when deoxyhemoglobin is transformed to oxyhemoglobin, is known as the alkaline Bohr effect.

The protons are derived from the partial dissociation of acid residues with pK_a values within ±1.5 pH units of the solution pH, that change from a higher to lower pK_a on the transformation of the T to R conformation of the hemoglobin. For example, His-146(β) in the deoxy conformation is predominantly in its imidazolium form (positively charged acid form), which is stabilized by a favorable interaction with the negatively charged side chain of Asp-94(β). This ion-pair interaction makes it more difficult to remove the imidazolium proton, and thus raises the pK_a of the imidazolium to a higher value than normally found for a free imidazolium ion in solution, where a stabilization by a proximate negatively charged group does not normally occur. However, on conversion of the protein to the R conformation, the strength of this ionic interaction is broken and the imidazolium assumes a lower pK_a. The decrease in the histidine's pK_a at blood pH results in the conversion of some of the acid form of the histidine to its conjugate base (imidazole) form, with the dissociation of free protons that form a part of the Bohr effect. Other acid groups in the protein contribute the additional protons due to analogous changes in their pK_a to lower values, correlated with the change in hemoglobin T to R conformation (Table 3.6).

The Bohr effect may fit the definition of an allosteric mechanism. An *allosteric mechanism* is a common process in protein molecules in

Table 3.6 Contributions of Particular Ionized Residues to the Alkaline Bohr Effect[a]

Group	*Deoxy* pK_a	*Oxy* pK_a	*Percent Contribution*
Val-1 (α) ($-NH_3^+$)[b]	8.0	7.33	39
Val-1 (β) ($-NH_3^+$)	7.03	6.78	1
His-117 (β)[b]	8.19	7.71	24
His-146 (β)[c]	8.43 (7.6)[d]	8.18 (6.7)[d]	16[c] (~50)[d]
Other residues			~20

Data From J. B. Matthew, G. I. H. Hanania, and F. R. N. Gurd, *Biochemistry*, **18**, 1928 (1979).

[a] Determined at pH 7.6 and at ionic strength of 0.1 M.

[b] Includes effect of Cl^- bound in deoxy hemoglobin at Val-1 (α) and His-117 (β).

[c] In this calculation by Matthew et al. rupture of His-146 (β)—Asp-94 (β) salt bridge on change from deoxy- to oxyhemoglobin is assumed not to occur.

[d] Found for solution conditions in which the His-146 (β)—Asp-94 (β) salt bridge is broken [I. M. Russu, N. T. Ho, and C. Ho, *Biochemistry* **19**, 1043 (1980)]; pK_a's in table are corrected for D_2O effect.

which substrate association is influenced by the binding of other molecules that are not direct substrates of the protein. In an allosteric process there must be separate binding sites on the protein for substrate (e.g., O_2 in the case of hemoglobin) and effector (inhibitor or activator) molecules that exert allosteric control. As the effector molecule's binding site by definition is distinctly separate from that of the substrate's binding site, the effector molecule acts to increase or decrease the affinity of the substrate at the substrate binding site by either causing a conformational change or stabilizing a particular conformation of the protein. With regard to the Bohr effect, it is evident that proton binding sites exist in the hemoglobin molecule to which the binding of protons, the effector "molecules," thermodynamically stabilizes and thus increases the concentration of the T form with respect to the R form. By increasing the ratio T/R on stabilizing the T conformation, the binding of protons to their effector sites is correlated with a poorer observed affinity of hemoglobin for oxygen.

The Bohr effect has important physiological consequences. Cells metabolizing at high rates, with high requirements for molecular oxygen, produce carbonic acid and lactic acid, which act to increase the hydrogen ion concentration in the cell's environment. As the increase in hydrogen ion concentration forces the equilibrium of equation (3.10) to the left, increased oxygen is thus dissociated from hemoglobin.

3.3 STRUCTURE AND MECHANISM OF ACTION OF ANTIBODY MOLECULES

Antibody molecules are immunoglobulins produced by an organism in response to the invasion of foreign compounds, such as proteins, carbohydrates, and nucleic acid polymers. The antibody molecule noncovalently associates to the foreign substance, initiating a process by which the foreign substance is eliminated from the organism.

Materials that elicit antibody production in an organism are called *antigens.* An antigen may contain multiple *antigenic determinants,* which are small regions of the antigen molecule that specifically elicit the production of antibody to which the antigen binds. In proteins, for example, an antigenic determinant may comprise only six or seven amino acids of the total protein.

A *hapten* is a small molecule that cannot alone elicit the produc-

tion of antibodies specific to it. However, when covalently attached to a larger molecule it can act as an antigenic determinant and elicit antibody synthesis. Whereas the hapten molecule requires attachment to a larger molecule to elicit the synthesis of antibody toward it, when detached from its carrier, the hapten will retain its ability to bind strongly to antibody.

It is estimated that a human individual can potentially produce at least 1×10^6 different antibody structures. However, all the antibodies have a similar basic structure. The elucidation of this structure has been accomplished to a great extent from studies of immunoglobulin primary structures. In addition, electron microscopy and recently obtained x-ray diffraction data have immeasurably added to our knowledge of the three-dimensional structure of the antibody molecule.

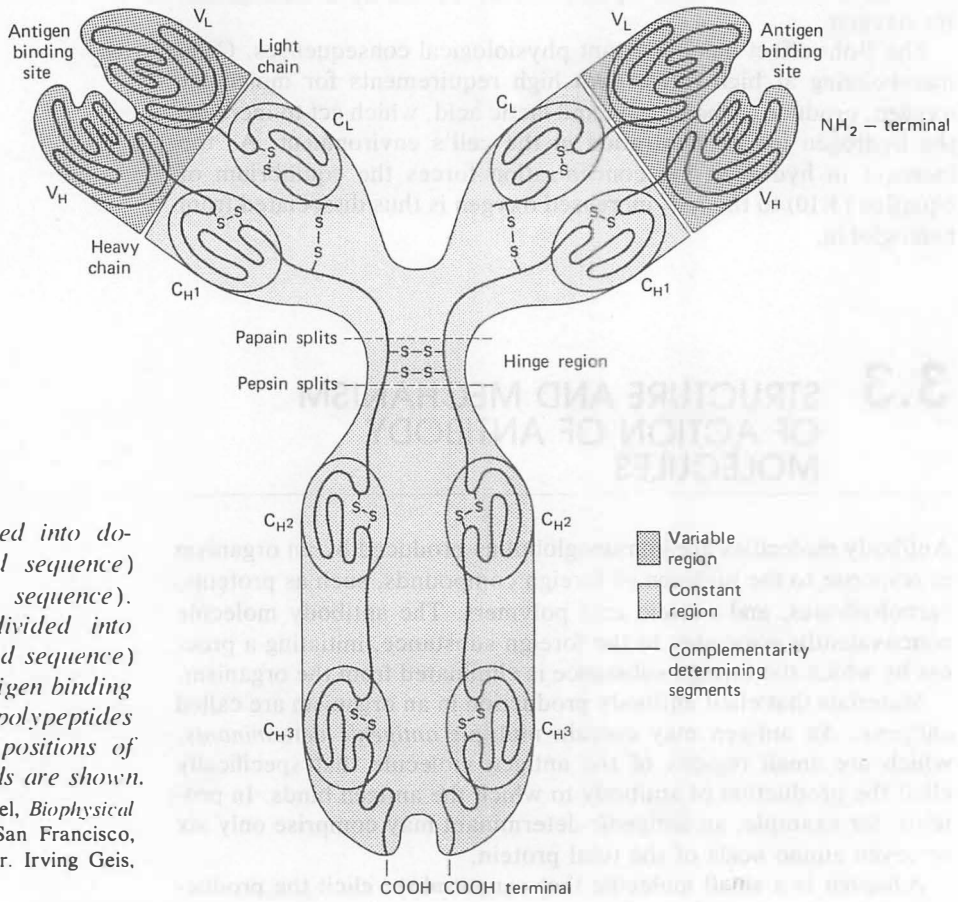


Figure 3.11
Diagrammatic structure for IgG.
The light chains (L) are divided into domains V_L (variable amino acid sequence) and C_L (constant amino acid sequence). The heavy chains (H) are divided into domains V_H (variable amino acid sequence) and C_H1, C_H2, and C_H3. The antigen binding sites are V_H–V_L. "Hinge" polypeptides interconnect the domains. The positions of inter- and intrachaincystine bonds are shown.
From C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, Part I, W. H. Freeman, San Francisco, 1980. Reprinted with permission of Mr. Irving Geis, N.Y.

Table 3.7 Immunoglobulin Classes

Classes of Immunoglobulin	*Approximate Mol Wt*	*H Chain Isotype*	*Carbohydrate by Weight (%)*	*Concentration in Serum (mg/100 ml)*
IgG	150,000	γ, 53,000	2–3	600–1800
IgA	170,000–720,000[a]	α, 64,000	7–12	90–420
IgD	160,000	δ, 58,000	—	0.3–40
IgE	190,000	ϵ, 75,000	10–12	0.01–0.10
IgM	950,000[a]	μ, 70,000	10–12	50–190

[a] Forms polymer structures of basic structural unit.

Structural studies of proteins require a pure homogeneous preparation of the protein. Such a sample of an antibody protein is extremely difficult to prepare because of the wide diversity of antibody molecules present in any one organism. Accordingly, immunoglobulins for structural studies have mostly been obtained from individuals with the disease of multiple myeloma. In this disease there is a loss of control of antibody synthesis in a single type of antibody-synthesizing cell from which large amounts of antibody of a single distinct structure is produced. This antibody is identical in its general structure to normal antibodies and can be easily purified in large amounts.

Primary Structure of Antibody (Immunoglobulin) Molecules

Antibody molecules are composed of four polypeptide chains, of which there are two chains of an identical primary structure and another two chains also with an identical primary structure. In the most common antibody, which is of the immunoglobulin G (IgG) class, the larger chain has approximately 440 amino acids (mol w 50,000). These chains are designated the *heavy chains,* H. The other kind of chain contains one-half the number of amino acids of the H chain, and are designated the *light chains,* L (mol w 25,000). The four polypeptide chains are covalently interconnected by disulfide cystine bonds, generating a single covalent structure for the antibody molecule (Figure 3.11).

In the other classes of immunoglobulins (see Table 3.7) the H chains have a slightly higher molecular weight than the antibodies of the IgG class. There is a small amount of carbohydrate (2 to 12%, depending on immunoglobulin class) attached to the H chain, and

CLIN. CORR. 3.1 THE COMPLEMENT PROTEINS

The complement proteins are composed of at least 11 distinct proteins in plasma. They are triggered by the association of IgG or IgM binding to antigens on the outer cell membrane of invading bacterial cells, protozoa, or tumor cells. On initiation by the immunoglobulin binding event, the 11 complement proteins become activated and associate with the cell membrane, causing a lysis of the membrane and cellular death.

Many of the complement proteins are precursors of proteolytic enzymes that circulate in plasma in a nonactive covalent form prior to activation. On their activation to active enzymes during the complementation process, they will in turn catalytically activate a succeeding protein of the pathway by facilitating the hydrolysis of a specific peptide bond in the second protein. The inactive precursor forms of enzymes are referred to as proenzymes or zymogens. The activation of enzymes by specific proteolysis (i.e., hydrolysis of a specific peptide bond in its primary structure) is an important general method for activating other extracellular protease enzymes. For example, the enzymes that catalyze blood clot formation, induce fibrinolysis of blood clots, and digest dietary proteins in the gut are all activated by a specific proteolysis catalyzed by a second enzyme.

The classical complementation reaction is initiated by the binding of IgG or IgM to cell surface antigens. The apparent exposure of a complement binding site in the antibody's F_c region on antigen association causes the binding of the C1 complement proteins, which is a protein complex composed of three individual proteins: C1q

(mol wt 400,000), C1r (mol wt 180,000), and C1s (mol wt 86,000). The C1r and C1s proteins become active enzymes on association with the immunoglobulin on the cell surface. The activated C1 complex ($C1^a$) catalytically hydrolyzes a peptide bond in C2 (mol wt 117,000) and in C4 (mol wt 206,000) complement proteins, which then form a complex that also associates to the cell surface. The now active C2–C4 complex has a catalytic activity that hydrolyzes a peptide bond in complement protein C3 (mol wt 180,000). Activated C3 protein binds to the cell surface, and the activated C2–C4–C3 complex activates protein C5 (mol wt 180,000). Activated protein C5 will associate with complement proteins C6 (mol wt 110,000), C7 (mol wt 100,000), C8 (mol wt 163,000), and six molecules of complement protein C9 (mol wt 79,000). This multiprotein complex binds to the cell surface and initiates membrane lysis.

The mechanism is a cascade type in which amplification of the trigger event occurs. Thus, activated C1 can activate multiple molecules of C4–C2–C3, and each activated C4–C2–C3 complex can in turn activate many molecules of C5–C9.

The series of reactions in the classical complementation pathway is summarized in the scheme below, where a and b designate the proteolytically modified proteins and a line above a protein indicates an enzyme activity.

$$\text{IgG or IgM} \xrightarrow{\text{Clq,Clr,Cls}} \overline{\text{C1a}} \xrightarrow{\text{C2,C4}}$$

$$\text{C4b} \cdot \overline{\text{C2a}} \xrightarrow{\text{C3}}$$

$$\text{C4b} \cdot \overline{\text{C2a}} \cdot \text{C3b} \xrightarrow{\text{C5,C6,C7,C8,6C9}}$$

$$\text{C5b} \cdot \text{C6} \cdot \text{C7} \cdot \text{C8} \cdot 6\text{C9}$$

There is an "alternative pathway" for C3 complement activation, initiated by aggregates of IgA or by bacterial polysaccharide in the absence of immunoglobulin binding to cell membrane antigens. This alternative pathway involves the proteins

thus the antibodies may be classified as glycoproteins (see Section 3.5).

In the three-dimensional antibody structure, each H chain is associated with an L chain, aligned so that the NH_2-terminal ends of both chains are near each other. The L and H chains then run in a general parallel direction with each other from their NH_2-terminal ends to their COOH-terminal ends. As the L chain is one-half the size of the H chain, only the NH_2-terminal one-half of the H chain (~214 amino acids in the IgG class) is associated with the L chain. The remaining COOH-terminal one-half of each of the two H chains within the antibody structure then align with each other. This is diagrammatically shown in Figure 3.11.

Constant and Variable Regions of Primary Structure

A comparison of the amino acid sequences of antibody molecules elicited toward different antigens shows that there are regions of exact homology and regions of high variability. In particular, the amino acid sequences of the NH_2-terminal one-half of the L chains and the NH_2-terminal one-quarter of the larger H chains are highly variable among the different antibody molecules. These NH_2-terminal segments of the L and H chain primary structures are designated the variable (V) regions because of the significant differences in amino acid sequence observed in these comparisons. Within the V region primary structure certain segments are observed to be even more variable than other segments. These segments are termed "hypervariable" regions. Three hypervariable regions of between 5 and 7 residues in the NH_2-terminal region of the L chain and four hypervariable regions of between 6 and 17 residues in the NH_2-terminal region of the H chain are commonly found.

In contrast, a comparison of the amino acid sequences from different antibodies shows that the COOH-terminal three-quarters of the H chains and the COOH-terminal one-half of the L chains are mostly constant in sequence. These regions of the polypeptide sequences are named the constant (C) regions of primary structure.

It is the C regions of the H chains that determine the class to which the antibody belongs (see below). In addition, the C region provides for the binding of complement proteins (Clin. Corr. 3.1) and provides the site necessary for antibodies to cross the placental membrane. The V regions determine the antigen specificity of the antibody molecule.

Different Classes of Immunoglobulins

The C regions of the H chains are homologous within a particular class of immunoglobulins, but differ significantly from the amino acid

sequence in the C regions of the H chains of other antibody classes. These differences in sequence within the C regions of the H chains are responsible for the different physical characteristics of the immunoglobulin classes.

In some cases the H chain promotes polymerization of the antibody molecules. Thus covalent $(LH)_2$ antibody units in the IgA class are sometimes found in dimer forms of the structure $[(LH)_2]_2$. The IgM antibodies are in a pentamer quaternary configuration of the basic $(LH)_2$ covalent structure, giving it the formula $[(LH)_2]_5$. The different H chains of distinct amino acid sequence are designated $\gamma, \alpha, \mu, \delta$, and ε for the H chains found in the IgG, IgA, IgM, IgD, and IgE immunoglobulin classes, respectively (Table 3.7; Clin. Corr. 3.2).

Although IgG is the major immunoglobulin in plasma, the biosynthesis of a specific IgG antibody in significant concentrations after exposure to a new antigen takes about 10 days (Clin. Corr. 3.3). In the absence of an initially high concentration of IgG to a specific antigen, the IgM class of antibodies, which are initially synthesized at faster rates, will associate with the antigen and thus serve as a first line of defense against the foreign antigen until the large quantities of IgG are produced (Figure 3.12; Clin. Corr. 3.3).

Two types of L chain constant sequences are synthesized, either of which are found combined with the five different classes of H chains. The two types of L chains are designated the *lambda* (λ) chain and the *kappa* (κ) chain.

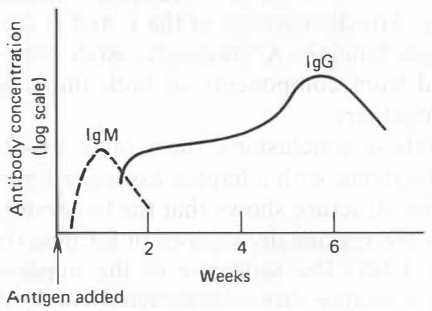


Figure 3.12
Time course of specific antibody IgM and IgG response to added antigen.
Based on figure in L. Stryer, *Biochemistry*, W. H. Freeman, San Francisco, 1975, p. 733.

properdin (mol wt 184,000), C3 proactivator convertase (mol wt 24,000), and C3 proactivator (mol wt 93,000).

CLIN. CORR. 3.2 FUNCTIONS OF THE DIFFERENT ANTIBODY CLASSES

The IgA class of immunoglobulins are primarily found in the extravascular secretions—bronchial, nasal, and intestinal mucus secretions; tears; milk; and colostrum. As such, these immunoglobulins are the initial defense against invading viral and bacterial antigens prior to their entry in plasma or other internal space.

The IgM class are primarily found in plasma. They are the first of the antibodies to act in significant quantity on the introduction of a foreign antigen into a host's plasma. IgM antibodies can promote phagocytosis of microorganisms by macrophage and polymorphonuclear leukocytes and are also potent activators of complement (see Clin. Corr. 3.1). IgM can be found in many external secretions, but at levels less than IgA.

The IgG class is found in high concentration in plasma. Its response to foreign antigens takes a longer period of time than that of IgM. However, at its maximum concentration it is present in significantly higher concentrations than the IgM antibodies. Like IgM antibodies, IgG antibodies can promote phagocytosis by phagocytic cells in plasma and can activate complement.

The normal biological functions of the IgD and IgE classes of immunoglobulins are not known. However, it is clear that the IgE antibodies play an important role in allergic responses, which are a cause of anaphylactic shock, hay fever, and asthma.

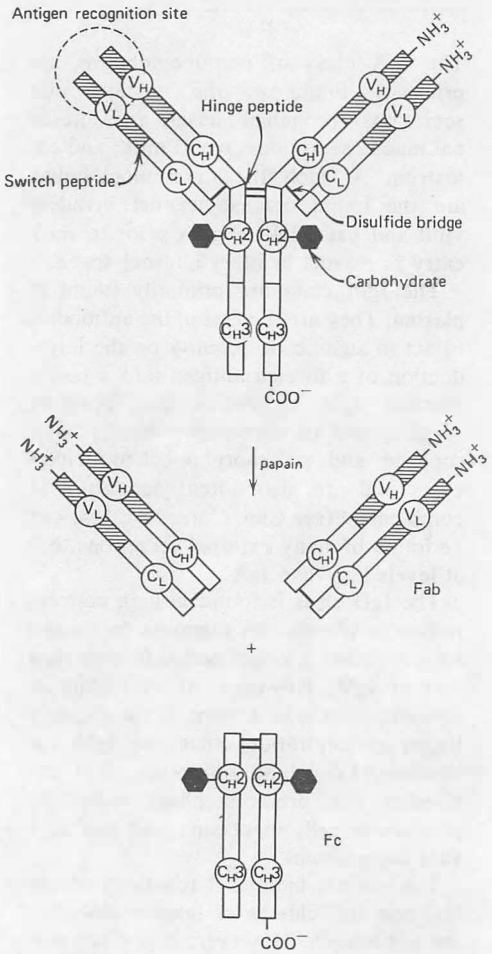


Figure 3.13
Hydrolysis of IgG into two Fab and one Fc fragments by papain, a proteolytic enzyme.

Existence of Two Antigen Binding Sites Per Antibody Molecule

The NH_2-terminal V regions of the L and H chain pairs comprise an antigen binding site. As the basic antibody structure contains two LH pairs, there are two antigen binding sites per antibody molecule. Clear evidence that an antigen binding site exists in the LH chain NH_2-terminal region was obtained by chemical techniques. In these experiments an antibody is enzymatically hydrolyzed by the proteolytic enzyme papain at a single peptide bond located in the hinge peptide of each H chain (see Figure 3.13). On papain hydrolysis of this peptide bond, the antibody molecule is cleaved into three products, two of which are identical and comprise the NH_2-terminal segments of the H chain associated with the full L chain (Figure 3.13). *These NH_2-terminal H·L fragments can bind antigens with a similar affinity as does the whole antibody molecule.* The two NH_2-terminal fragments are designated the Fab (antigen binding) fragments. The second type of product from the papain hydrolysis is the COOH-terminal one-half of the H chains formed in a single covalent fragment designated the Fc (crystallizable) fragment. The Fc fragment cannot bind antigen. It is thus clear that there are two antigen binding sites per antibody molecule present in the NH_2-terminal H·L fragments, which comprise the variable (V) amino acid sequences of each of the chains.

The valency of 2 for each antibody molecule allows each antibody to bind to two different antigens. This property of bivalency facilitates the agglutination and precipitation of antigen molecules by allowing the antibodies to form an interconnected matrix of antigens and antibodies.

The L chain can be dissociated from its H chain segment within the Fab fragment by oxidation of disulfide bonds, followed by chromatography. The dissociation of the L and H chains in this way eliminates antigen binding. Accordingly, each antigen binding site must be formed from components of both the L and H variable regions acting together.

In support of these conclusions, the x-ray crystallographic structure of an Fab fragment with a hapten associated has been obtained (Figure 3.14). The structure shows that the hypervariable sequences of the V regions are specifically utilized in forming the antigen binding site (Figure 3.14). The sequence of the hypervariable regions apparently give a unique three-dimensional conformation for each antibody that makes it specific to the antigenic determinant with which it associates.

The strength of association between antibody and antigen is due to noncovalent forces (see Chapter 2). The complementarity of the

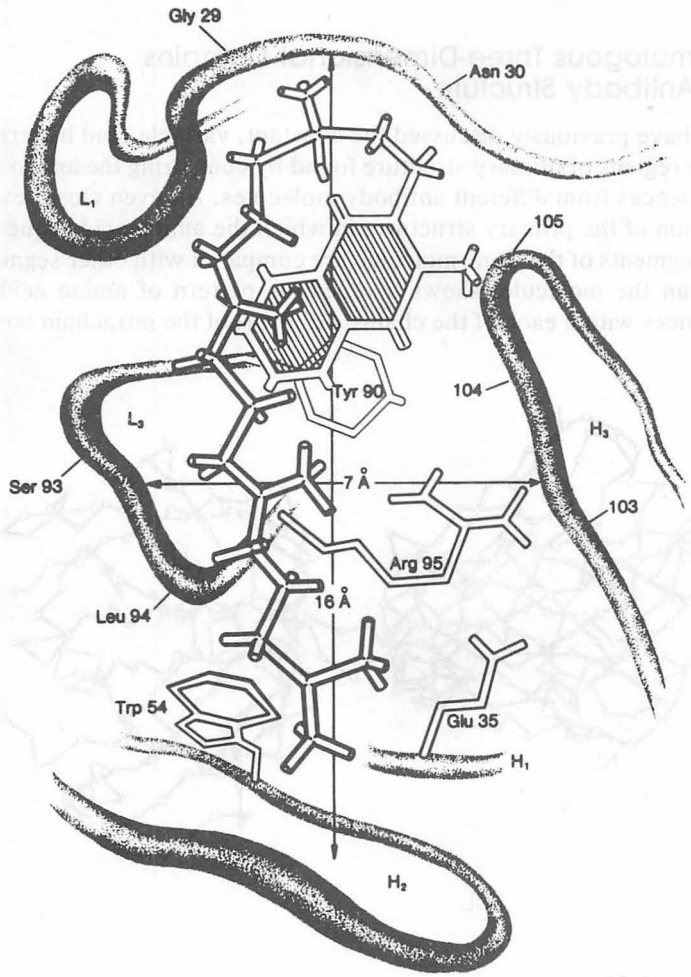


Figure 3.14
Structure of antigen (vitamin K_1—OH) bound to variable region of antibody.
Hypervariable regions designated L_1 and L_3 in light chains and H_1, H_2, and H_3 in heavy chains are shown in van der Waals contact with antigen.

Reprinted with permission from L. M. Amzel, R. J. Poljack, F. Saul, J. M. Varga, and F. F. Richards, *Proc. Natl. Acad. Sci., USA*, 71, 1427, 1974.

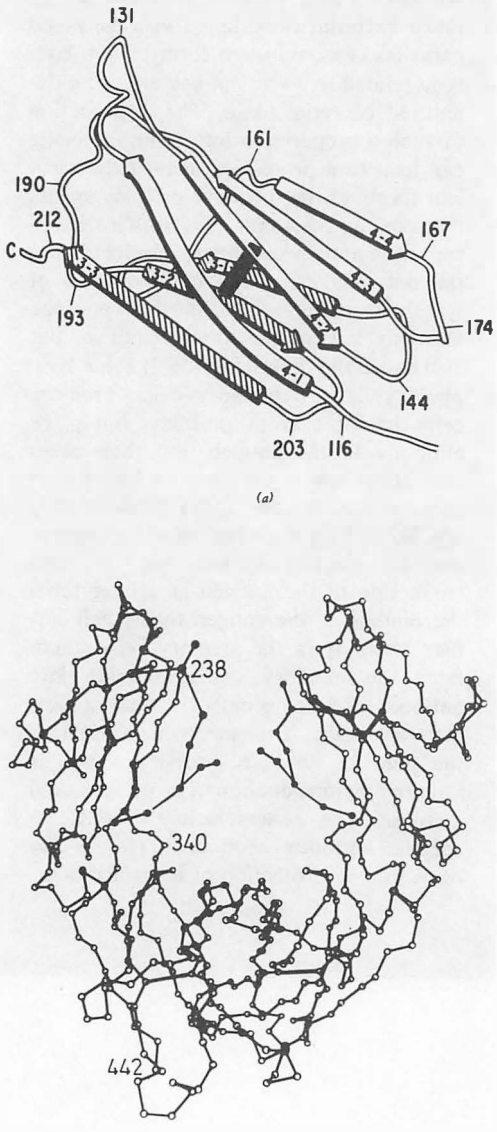
CLIN. CORR. 3.3 IMMUNIZATION

An immunizing vaccine may consist of killed bacterial cells, killed viruses, killed parasites, a nonvirulent form of live bacteria related to a virulent bacteria, or a denatured bacterial toxin. The introduction of such a preparation into a human being can lead to a protection against the virulent forms of the infective or toxic agents that contain the same antigen. This is because the antigens in the nonvirulent material not only cause the differentiation of lymphoid cells into cells that produce antibody toward the foreign antigen, but also cause the differentiation of some lymphoid cells into memory cells. Memory cells do not excrete antibody, but place antibody to the antigen into their outer membrane where they act as future sensors for the antigen. These memory cells are like a long-standing radar for the potentially virulent antigen. On the reintroduction of the antigen at a later time, the binding of the antigen to the cell surface antibody in the memory cells stimulates the memory cell to divide into antibody-producing cells as well as new memory cells. This mechanism reduces the time for antibody production that is required on introduction of an antigen, and increases the concentration of antigen-specific antibody produced. This is the basis for the protection of immunization.

structures of the antigenic determinant and antigen binding site within the antibody results in extremely high equilibrium affinity constants, between 10^5 and 10^{10} M^{-1} (strength of 7–14 kcal/mol) for this noncovalent association.

Homologous Three-Dimensional Domains of Antibody Structure

We have previously discussed the constant, variable, and hypervariable regions of primary structure found by comparing the amino acid sequences from different antibody molecules. An even closer examination of the primary structure, in which the amino acid sequences of segments of the same molecule are compared with other segments within the molecule, shows a repeating pattern of amino acid sequences within each of the chains. Analysis of the intrachain homol-



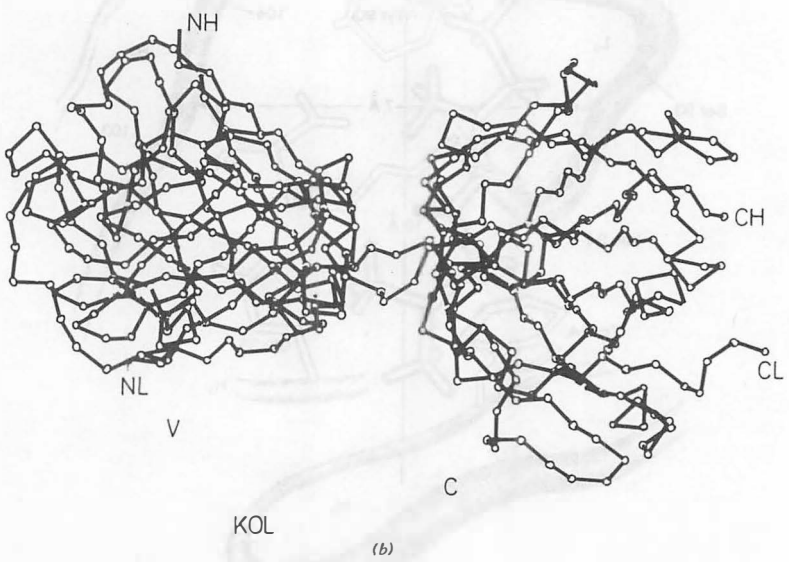


Figure 3.15

(a) Schematic diagram of folding of a C_L domain, showing β-pleated sheet structure. (b) The α-carbon (○) structure of the Fab fragment of IgG Kol showing the V_L–V_C and C_H1–C_L domains interconnected by hinge polypeptide, (c) Fc fragment from a human immunoglobin showing C_H2 and C_H3 domain.

Figure *a* from Edmundson *et al.*, *Biochemistry* **14**, 3953 (1975). Copyright 1975 American Chemical Society. Figures *b* and *c* from R. Huber, J. Deisenhofer, P. M. Coleman, M. Matsushima, and W. Palm, in *The Immune System*. 27th Mosbach Colloquium, Springer-Verlag, Berlin, 1976, p. 26. Reprinted with permission.

ogy indicates that approximately 110 amino acids form a primary structure unit that is repeated twice in an L chain and four times in an H chain (see Figure 3.11). Each of these repeating sequences or units are cyclized by an intrachain cystine formed from two cysteines near the two ends of this sequence unit. The periodic similarities in sequence, even into the NH_2-terminal V regions, suggests a mechanism of antibody evolution in which current antibody genes have evolved by gene duplication of a primordial gene that coded for a protein 110 amino acids in length. Mutations in the duplicated DNA sequence over time resulted in the different but analogous regions seen in the present antibody protein.

Based on the presence of analogous regions in the amino acid sequence, it was proposed that each of the analogous segments may have a similar tertiary conformation. X-ray diffraction studies confirm this contention. The x-ray data dramatically show periodic domains in the H and L chains with a similar three-dimensional conformation as predicted from the study of the primary structure. The globular domains are formed from the winding of the polypeptide chains back and forth upon itself forming two β-pleated sheet structures that run roughly parallel to each other. In three dimensions the two sheets form a structure that appears like a double blanket with a bulky hydrophobic core (Figure 3.15). In the NH_2-terminal domain a shallow crevice is opened into the center of the hydrophobic core that binds to antigen. The hypervariable sequences are present in this NH_2-terminal domain crevice, standing out as different in conformation from the rest of the folded domain.

The x-ray structure of isolated crystalline Fab fragments, Fc fragments, and intact immunoglobulin molecules show a variability in the relative arrangements between domains due to rotational freedom about the hinge region polypeptide segments that interconnect the domains (Figure 3.16). Furthermore, the structures suggest that on antigen binding conformational changes of the binding domain cause new interdomain relationships that "open" functional sites in the constant domains of the H chain that, for example, can facilitate the binding of the complement proteins on antigen association to the immunoglobulin (see Clin. Corr. 3.1).

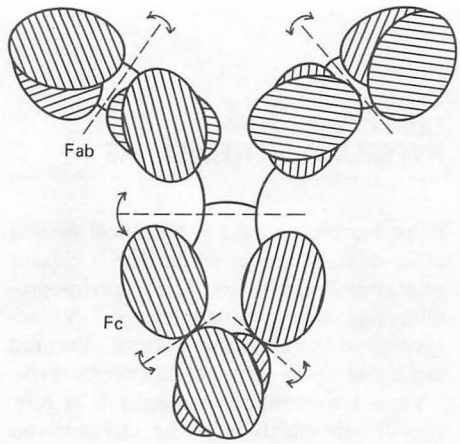


Figure 3.16
Rotations of domain regions with respect to each other about hinge polypeptides.
Reprinted with permission from R. Huber, *Trends Biochem. Sci.*, **4**, 270 (1979).

3.4 PLASMA LIPOPROTEINS

A lipoprotein is a multicomponent complex of protein and lipids of characteristic density, molecular weight, size, and chemical com-

CLIN. CORR. 3.4 HYPERLIPOPROTEINEMIAS

Hyperlipoproteinemia is a clinical finding of an overabundance of particular classes of plasma lipoproteins. The hyperlipoproteinemias are classified (types I–V) according to the abnormal pattern observed in plasma lipoprotein gel electrophoresis.

Type I hyperlipoproteinemia is a rare disorder in which only the chylomicron fraction is abnormally increased in gel electrophoresis. This type of hyperlipoproteinemia appears in patients with a genetic deficiency of the enzyme lipoprotein lipase in adipose tissue. These patients cannot metabolize lipid normally, and increased amounts of fat are found in the blood with a corresponding higher level of chylomicrons. This condition leads to hepatosplenomegaly.

Type II hyperlipoproteinemia (also referred to as hyper-β-lipoproteinemia or hypercholesterolemia) is characterized by an increased β-lipoprotein (LDL; type IIA) or both an increased β-lipoprotein and pre-β-lipoprotein (VLDL and LDL; type IIB) bands in plasma electrophoresis. These types of hyperlipoproteinemia exist in both a hereditary form and in an acquired form. Patients with type II hyperlipoproteinemia have an increased risk of atherosclerosis, especially of the coronary arteries, and they have high amounts of cholesterol in plasma.

Type III hyperlipoproteinemia is a rare disorder that is characterized by the appearance of an abnormal lipoprotein in plasma, which is referred to as "floating β-lipoprotein." This abnormal lipoprotein appears as a broad band between the normal β-lipoprotein and pre-β-lipoprotein in

position. These complexes of protein and lipids are held together by noncovalent forces. While a certain typical chemical composition and molecular weight exists for each type of lipoprotein complex, there may exist no *exact* stoichiometry among the components of the complex.

Lipoprotein complexes serve a wide variety of functions in cellular membranes and in the transport and metabolism of lipids. In plasma, for example, lipoproteins transport lipids from the sites of their absorption into blood to the various tissues of the organism that utilize lipids. The plasma lipoproteins are the most extensively characterized of the lipoproteins, and it is the example of the plasma lipoproteins that will be described in this section.

Classification of the Plasma Lipoproteins

The classification of the plasma lipoproteins is difficult as the physical and chemical characteristics of these complexes are often heterogeneous. However, the most popular system for classification of plasma lipoprotein particles is based on the criterion of density.

Four hydrated density classes of plasma lipoproteins are now widely recognized in normal fasting humans (Table 3.8). These are the high density lipoproteins (HDL) (d = 1.063–1.210), the low density lipoproteins (LDL) (d = 1.019–1.063), the intermediate density lipoproteins (IDL or LDL_2) (d = 1.006–1.019), and the very low density lipoproteins (VLDL) (d = 0.95–1.006) (see Clin. Corr. 3.4). In addition, lipid particles with small amounts of protein that are less dense than the VLDL appear in plasma after a fatty meal. These are the chylomicrons with $d < 0.95$.

The density of a lipoprotein particle is determined by the density

Table 3.8 Hydrated Density Classes of Plasma Lipoproteins

Lipoprotein Fraction	*Density (g/ml)*	*Flotation Rate (S_f) (Svedberg Units)*	*Molecular Weight*	*Particle Diam (Å)*
HDL	1.063–1.210	—	HDL_2, 4×10^5	70–130
			HDL_3, 2×10^5	50–100
LDL (or LDL_2)	1.019–1.063	0–12	2×10^6	200–280
IDL (or LDL_1)	1.006–1.019	12–20	4.5×10^6	250
VLDL	0.95–1.006	20–400	5×10^6–10^7	250–750
Chylomicrons	<0.95	>400	10^9–10^{10}	10^3–10^4

Data from A. K. Soutar and N. B. Myant, in *Chemistry of Macromolecules*, IIB, R. E. Offord, ed., University Park Press, Baltimore, Md., 1979.

of the solution into which the lipoprotein floats in an ultracentrifugation experiment or from its flotation rate. The flotation rate for a lipoprotein particle under standard conditions (NaCl solution $d = 1.063$) is denoted by the term S_f (Table 3.8).

The plasma lipoproteins migrate with the α- and β-globulin fractions in plasma electrophoresis (Figure 3.17).

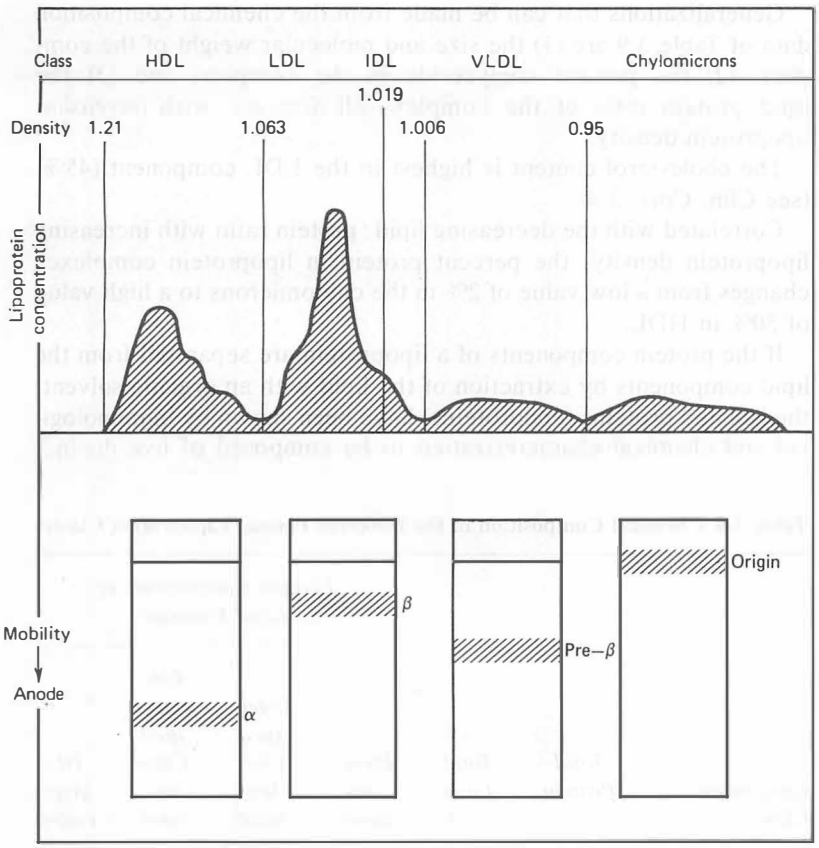


Figure 3.17
Correspondence of lipoprotein density classes with electrophoresis mobility in a plasma electrophoresis.
In the upper diagram the ultracentrifugation schlieren pattern is shown. At the bottom, electrophoresis on a paper support shows the mobilities of major lipoprotein classes with respect to the α and β electrophoresis bonds.

Reprinted with permission from A. K. Soutar and N. B. Myant, in *Chemistry of Macromolecules*, IIB, R. E. Offord, ed., University Park Press, Baltimore, Md., 1979.

plasma electrophoresis. The disease has a hereditary form, although type III sometimes also appears as a secondary effect to hypothyroidism. Patients with type III hyperlipoproteinemia show the early appearance of atherosclerosis and have an increased risk of occlusive vascular diseases.

Type IV hyperlipoproteinemia (also referred to as pre-β-hyperlipoproteinemia, hypertriglyceridemia) is characterized by an increased pre-β-lipoprotein band in electrophoresis due to increased amounts of VLDL. It is thought that this type of hyperlipoproteinemia can exist in a hereditary form. Type IV is often found in conjunction with overweight in an individual characterized by larger than normal size fat cells in the adipose tissue. Type IV is often found in patients with a hereditary tendency toward diabetes mellitus. A reversible form of type IV is often found in individuals who imbibe excessive amounts of alcohol. Patients with type IV hyperlipoproteinemia have a high incidence of coronary artery disease and peripheral vascular disease.

Type V hyperlipoproteinemia is a rare disease that is sometimes expressed secondarily to diabetes mellitus, chronic pancreatitis, hepatopathy, and nephropathy. It is characterized by an increased chylomicron and pre-β-lipoprotein bands in electrophoresis.

An abnormal lipoprotein, designated *lipoprotein X*, that electrophoreses with the β-lipoprotein band and has a density of 1.040–1.045, occurs in patients with a deficiency of the enzyme lecithin : cholesterol acyltransferase (LCAT) and in patients with cholestasis. The apolipoprotein of lipoprotein X consists primarily of albumin with one or two of the ApoC proteins.

There is epidemiological evidence that human coronary disease has a lower incidence in individuals with higher relative levels of HDL or α-lipoproteins.

Composition of the Plasma Lipoproteins

The lipid fraction of the plasma lipoproteins contains significant amounts of triglycerides, phospholipids, free cholesterol, and cholesterol esterified with long-chain fatty acids, with other lipids present in small amounts (Table 3.9). The percent lipid content within each of the density classes is variable within extremes that set limits for each of the classes.

Generalizations that can be made from the chemical composition data of Table 3.9 are (1) the size and molecular weight of the complex, (2) the percent triglyceride in the complex, and (3) the lipid : protein ratio of the complex—all *decrease* with *increasing* lipoprotein density.

The cholesterol content is highest in the LDL component (45%) (see Clin. Corr. 3.4).

Correlated with the decreasing lipid : protein ratio with increasing lipoprotein density, the percent protein in lipoprotein complexes changes from a low value of 2% in the chylomicrons to a high value of 50% in HDL.

If the protein components of a lipoprotein are separated from the lipid components by extraction of the lipid with an organic solvent, the isolated proteins (apoproteins) have been shown by immunological and chemical characterization to be composed of five distinct

Table 3.9 Chemical Composition of the Different Plasma Lipoprotein Classes

			Percent Composition of Lipid Fraction			
Lipoprotein Class	*Total Protein (%)*	*Total Lipid (%)*	*Phospholipids*	*Esterified Cholesterol*	*Unesterified Cholesterol*	*Triglycerides*
HDL_2[a]	40–45	55	35	12	4	5
HDL_3[a]	50–55	50	20–25	12	3–4	3
LDL	20–25	75–80	15–20	35–40	7–10	7–10
IDL	15–20	80–85	22	22	8	30
VLDL	5–10	90–95	15–20	10–15	5–10	50–65
Chylomicrons	1.5–2.5	97–99	7–9	3–5	1–3	84–89

Data from A. K. Soutar and N. B. Myant, in *Chemistry of Macromolecules*, IIB, R. E. Offord, ed., University Park Press, Baltimore, Md., 1979.

[a] Subclasses of HDL.

Table 3.10 Apoprotein of the Human Plasma Lipoproteins (Values in Percent of Total Protein Present)[a]

Apolipoprotein	*HDL_2*	*HDL_3*	*LDL*	*IDL*	*VLDL*	*Chylomicrons*
ApoA-I	85	70–75	Trace	0	0–3	0–3
ApoA-II	5	20	Trace	0	0–0.5	0–1.5
ApoD	0	1–2	—	—	0	1
ApoB	0–2	0	95–100	50–60	40–50	20–22
ApoC-I	1–2	1–2	0–5	<1	5	5–10
ApoC-II	1	1	0.5	2.5	10	15
ApoC-III	2–3	2–3	0–5	17	20–25	40
ApoE	Trace	3–5	0	15–20	5–10	5

Data from A. K. Soutar and N. B. Myant, in *Chemistry of Macromolecules*, IIB, R. E. Offord, ed., University Park Press, Baltimore, Md., 1979.

[a] Values show variability from different laboratories.

types (Table 3.10). In some of the apoprotein types only a single distinct protein is currently known, while other types are composed of several distinct proteins.

The initially discovered apolipoproteins from the plasma lipoproteins were named apolipoprotein A (ApoA) for the major protein of the HDL fraction, and apolipoprotein B (ApoB) for the major protein isolated from the LDL fraction. ApoB is also the major protein in the IDL and VLDL fractions (Table 3.10). A third apolipoprotein (ApoC) is found in the IDL and VLDL. Other chemically and immunologically distinct apolipoproteins are ApoD and ApoE. ApoD is found in the HDL, where it is a minor component. This protein is sometimes designated $ApoA_3$, based on a common A designation for proteins isolated from the HDLs. (Clin. Corr. 3.5).

Within the A and C classes of apolipoproteins several distinct proteins have been identified. These proteins are distinguished either by their COOH-terminal amino acid residue or by a consecutive numbering system [i.e., A-I, A-II, C-I, C-II, etc. (Table 3.10)]. These polypeptide chains are present in defined ratios within each apolipoprotein. For example, ApoA-I and ApoA-II are normally present in HDL_3 in an approximate ratio of 3:1.

The molecular weights of the apolipoproteins vary from 6,000 (ApoC-I) to 33,000 (ApoE). ApoB is relatively insoluble even in detergents, and reports on its molecular weight are highly variable due to aggregation of the protein.

The apolipoproteins are found to have significant amounts of

CLIN. CORR. 3.5 HYPOLIPOPROTEINEMIAS

A-β-lipoproteinemia is a genetically transmitted disease that is characterized by an absence of chylomicrons, VLDL, and LDL due to an inability to synthesize apoprotein B.

Tangier's disease is a rare autosomal recessive inherited disease in which the HDL is 1 to 5% of its normal value. The clinical features are due to the accumulation of cholesterol in the lymphoreticular system, which may lead to hepatomegaly and splenomegaly. In this disease the plasma cholesterol and phospholipids are greatly reduced.

Deficiency of the enzyme lecithin: cholesterol acyltransferase is a rare disease that results in the production of lipoprotein X (see Clin. Corr. 3.4). Also characteristic of this disease is the decrease in the α-lipoprotein and pre-β-lipoprotein bands, with the increase in the β-lipoprotein (lipoprotein X) in electrophoresis.

α-helical secondary structure (Table 3.11), as determined by spectrophotometric methods (see Chapter 2).

Structure of the Lipoprotein Molecule

The diameter range for HDL is 95–130 Å and for LDL, 230–280 Å. The other density classes are less homogeneous in their molecular size, and their diameters are accordingly variable.

The structure of lipoprotein molecules have been investigated with a wide range of methods, including electron microscopy, x-ray diffraction scattering, chemical and enzymatic modification, and spectrophotometric techniques. In addition, calculations based on the physical properties of the individual components in a lipoprotein have led to significant insights into their structure. However, these

Table 3.11 Physical Characteristics of Plasma Apolipoproteins

Apolipoprotein	*Molecular Weight*	*Number of Amino Acids*	*Calc. Secondary Structure (Lipid-free)*
ApoA-I	28,300	243	55% α helix (~70% α helix in lipoprotein) ~10% β sheet ~35% disordered
ApoA-II	17,380	Dimer of two 77 amino acid chains	35% α helix
ApoB	8,000–300,000	—	In lipoprotein: ~25% α helix 37% β sheet 37% disordered
ApoC-I	6,600	57	55% α helix
ApoC-II	8,800	78	23% α helix
ApoC-III	8,700	79	22% α helix
ApoD	20,000	—	—
ApoE	33,000	—	?66%

Data from results summarized by A. K. Soutar and N. B. Myant, in *Chemistry of Macromolecules* IIB, R. E. Offord, ed., University Park Press, Baltimore, Md., 1979; J. D. Morrisett, R. L. Jackson, and A. M. Gotto, Jr., *Annu. Rev. Biochem.*, 44, 183 (1975); and L. C. Smith, H. J. Pownall, and A. M. Gotto, Jr., *Annu. Rev. Biochem.*, 47, 751 (1978).

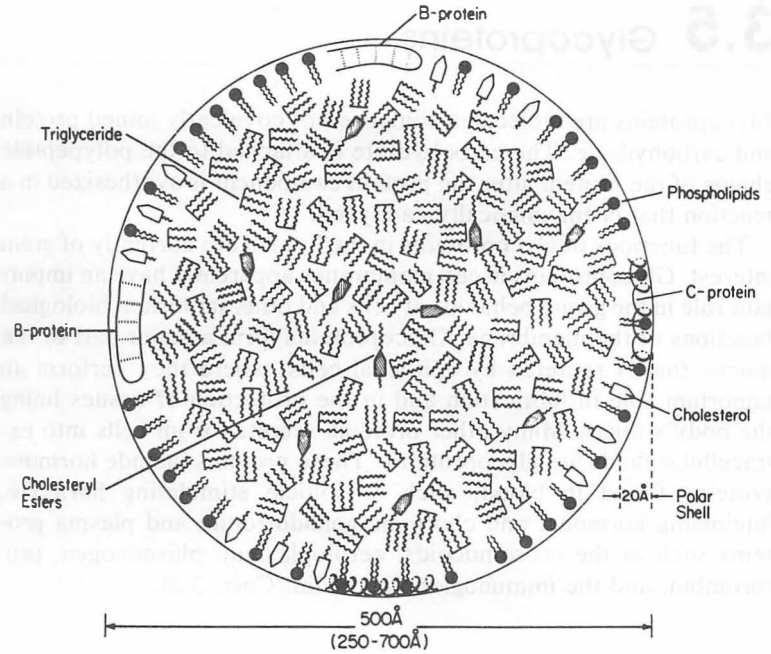


Figure 3.18
Generalized structure for the lipoproteins.
This structure is drawn for VLDL. Note that protein, phospholipids, and cholesterol (acid) form a shell of 20 Å on the outside.

Reprinted with permission from I. R. Morrisett, R. L. Jackson, and A. M. Gotto, *Biochim. Biophys. Acta*, **472**, 93 (1977).

studies have not yet yielded a definitive structure for plasma lipoproteins, and the structure discussed below may be considered speculative.

Calculations based on the volume in space occupied by the chemical components of a lipoprotein molecular complex show that the sum of the volume of the protein *plus* amphoteric lipid (e.g., phosphatidylcholine and cholesterol) components *in any of the plasma lipoprotein density classes* is just sufficient to occupy an outer shell 21.5 Å thick for the lipoprotein complex. The calculation thereby supports a structure of the lipoprotein in which the amphoteric components, which include the protein components, phosphatidylcholine and unesterified cholesterol, are on the outside of the complex forming an outer sphere approximately 21.5 Å thick, with their apolar parts facing inside and their polar and charged parts facing outside toward the water solvent. The nonpolar neutral lipids are inside the complex in contact with the nonpolar ends of the amphoteric molecules of the outer shell (Figure 3.18). The calculations show that this model is possible for complexes in all density classes, irrespective of their particle size. Furthermore, the model nicely explains the decrease in neutral lipid to protein percentage with changes in complex size (Table 3.9). The data from x-ray scattering, electron microscopy, and other techniques generally support this overall structure. More detailed, but currently speculative, structural models for plasma lipoproteins have been put forth based on the current experimental data.

CLIN. CORR. 3.6 FUNCTIONS OF GLYCOPROTEINS

Glycoproteins participate in a large number of normal and disease-related functions of clinical relevance. For example, many of the proteins in the outer cellular membranes are glycoproteins. Such proteins within the cell surface are antigens, which determine the blood antigen system (A, B, O), and the histocompatibility and transplantation determinants of an individual. Immunoglobulin antigenic sites and viral and hormone receptor sites in cellular membranes are often glycoproteins. In addition, the carbohydrate portions of glycoproteins in membranes provide a surface code for cellular identification by other cells and for contact inhibition in the regulation of cell growth. As such, changes in the membrane glycoproteins can be correlated with tumorgenesis and malignant transformation in cancer.

Most of the important plasma proteins, except albumin, are glycoproteins. These plasma proteins include the blood-clotting proteins, the immunoglobulins, and many of the complement proteins. Some of the protein hormones, such as follicle-stimulating hormone (FSH) and thyroid-stimulating hormone (TSH), are glycoproteins. The important structural protein collagen contains carbohydrate. The proteins found in the mucus secretions are carbohydrate-containing proteins, where they perform a role in lubrication and in the protection of epithelial tissue. The antiviral protein interferon is a glycoprotein.

3.5 Glycoproteins

Glycoproteins are molecules composed of covalently joined protein and carbohydrate. The carbohydrate is attached to the polypeptide chains of the protein after the protein component is synthesized in a reaction that is enzymatically catalyzed.

The functions of glycoproteins in the human are currently of great interest. Glycoproteins in cell membranes apparently have an important role in the group behavior of cells and other important biological functions of the membrane. Glycoproteins form a major part of the mucus that is secreted by epithelial cells, where they perform an important role in lubrication and in the protection of tissues lining the body's ducts. Many other proteins secreted from cells into extracellular fluids are glycoproteins. These proteins include hormone proteins found in blood, such as follicle stimulating hormone, luteinizing hormone, and chorionic gonadotropin; and plasma proteins such as the orosomucoids, ceruloplasmin, plasminogen, prothrombin, and the immunoglobulins (Clin. Corr. 3.6).

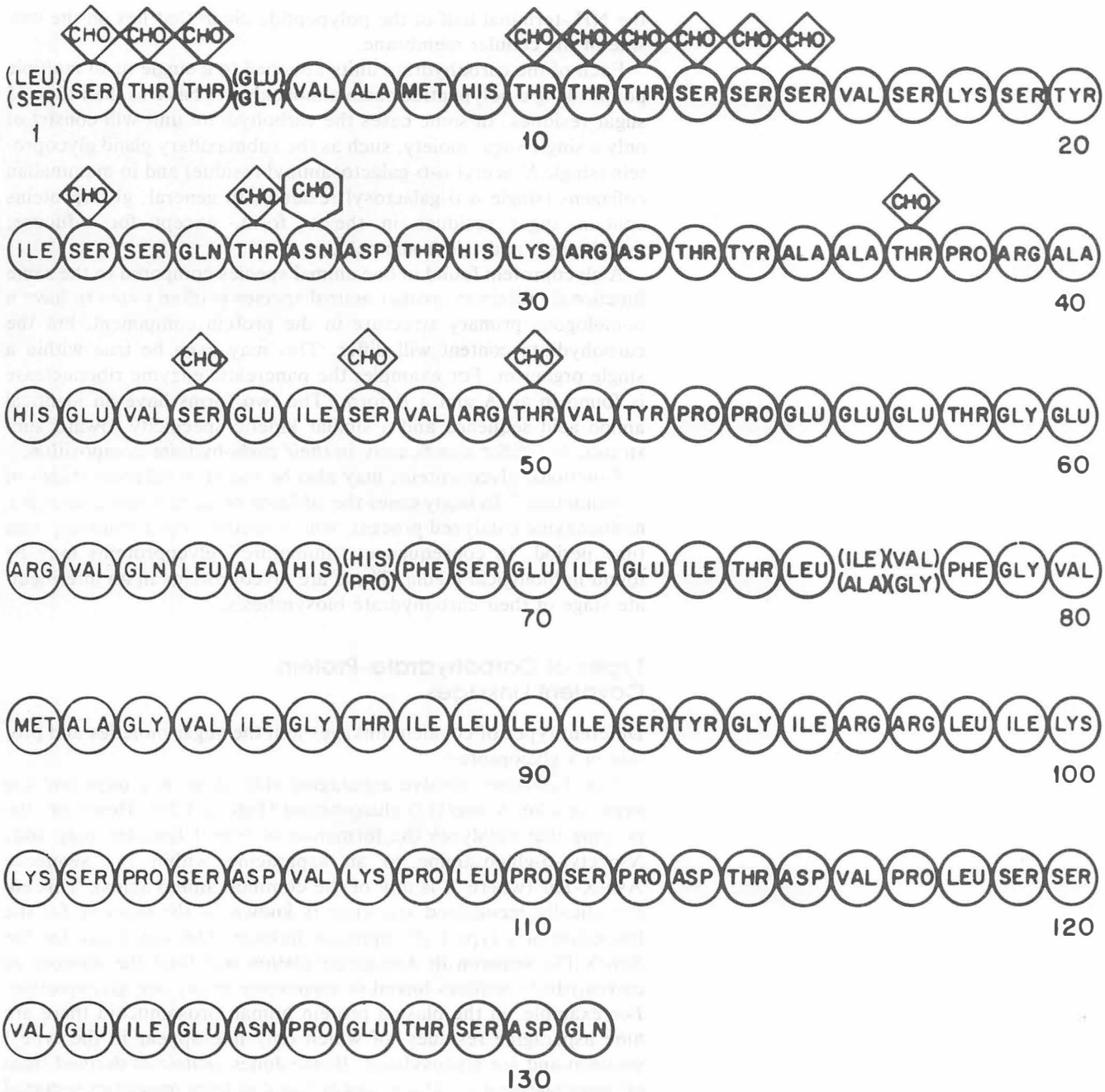
Carbohydrate Composition of Glycoproteins

The percent of carbohydrate within the glycoproteins is highly variable. Some glycoproteins such as IgG contain low amounts of carbohydrate (4%). A cell membrane glycoprotein, human glycophorin, has been found to contain 60% carbohydrate (Figure 3.19). Human ovary cyst glycoprotein is composed of 70% carbohydrate, and human gastric glycoprotein is 82% carbohydrate.

The carbohydrate can be distributed fairly evenly along the polypeptide chain of the protein component or concentrated into defined regions of the polypeptide chain of the protein component. For example, in human glycophorin A the carbohydrate is found in

→

Figure 3.19
Primary structure of human red blood cell membrane glycophorin A. *O-Glycosidic carbohydrate bonds to Ser and Thr are indicated by diamonds and N-glycosidic bonds to Asn by the hexagon. Residues 71–91 lie in the membrane; the NH_2-terminal end with carbohydrate attached is on the outside of the cell (residues 1–70) and the COOH-terminal end (residues 91–131) on the inside of the cell.*
Reprinted with permission from M. Tomita and V. T. Marchesi, *Proc. Natl. Acad. Sci., USA*, 72, 2964, 1975.


CHO CHO CHO
(LEU) (SER) SER THR THR (GLU) (GLY) VAL ALA MET HIS
CHO CHO CHO CHO CHO CHO
THR THR THR SER SER SER VAL SER LYS SER TYR
1
10
20
CHO
CHO
CHO
ILE SER SER GLN THR ASN ASP THR HIS LYS ARG ASP THR TYR ALA ALA
CHO
THR PRO ARG ALA
30
40
CHO
CHO
CHO
HIS GLU VAL SER GLU ILE SER VAL ARG THR VAL TYR PRO PRO GLU GLU GLU THR GLY GLU
50
60
ARG VAL GLN LEU ALA HIS (HIS) (PRO) PHE SER GLU ILE GLU ILE THR LEU (ILE) (VAL) (ALA) (GLY) PHE GLY VAL
70
80
MET ALA GLY VAL ILE GLY THR ILE LEU LEU ILE SER TYR GLY ILE ARG ARG LEU ILE LYS
90
100
LYS SER PRO SER ASP VAL LYS PRO LEU PRO SER PRO ASP THR ASP VAL PRO LEU SER SER
110
120
VAL GLU ILE GLU ASN PRO GLU THR SER ASP GLN
130

the NH_2-terminal half of the polypeptide chain that lies on the outside of the cellular membrane.

Each of the carbohydrate units attached to a single or to multiple points along a polypeptide chain almost always has less than 12 to 15 sugar residues. In some cases the carbohydrate unit will consist of only a single sugar moiety, such as the submaxillary gland glycoprotein (single *N*-acetyl-α-D-galactosaminyl residue) and in mammalian collagens (single α-D-galactosyl residue). In general, glycoproteins contain sugar residues in the D form, except for L-fucose, L-arabinose, and L-iduronic acid.

A glycoprotein found in one animal species compared to the same functional protein in another animal species is often found to have a homologous primary structure in the protein component, but the carbohydrate content will differ. This may even be true within a single organism. For example, the pancreatic enzyme ribonuclease is found in an A and a B form. The two forms have an identical amino acid sequence and a similar kinetic specificity toward substrates, but differ significantly in their carbohydrate composition.

Functional glycoproteins may also be found in different stages of "completion." In many cases the addition of carbohydrate units is a multienzyme-catalyzed process, which occurs over a relatively long time period. In consequence, "immature" glycoproteins may be found in biological media. These are glycoproteins in an intermediate stage of their carbohydrate biosynthesis.

Types of Carbohydrate–Protein Covalent Linkages

Different types of covalent linkages join the sugar moieties and protein in a glycoprotein.

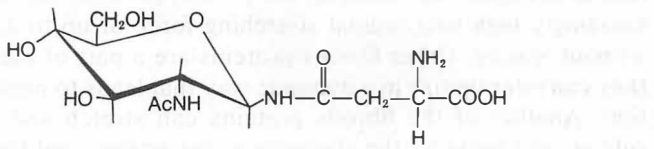
Type I linkages involve asparagine side chain R groups and the sugar residue *N*-acetyl-D-glucosamine (Figure 3.20). However, the enzyme that catalyzes the formation of type I linkages only adds *N*-acetyl-D-glucosamine to an asparagine within the sequence Asn-X-Thr (where X is any of the common animo acids). This enzymatically recognized sequence is known as the *sequeon* for the formation of a type I glycoprotein linkage. The specificity for the Asn-X-Thr sequeon in Asn glycosylation will limit the number of carbohydrate residues linked to asparagine in any one glycoprotein. For example, in the plasma protein human orosomucoid there are nine asparagine residues, of which only five appear in the type I sequeon and are glycosylated. Bence-Jones protein is derived from an immunoglobulin L chain, and is found in large quantities secreted in the urine of patients with malignant myeloma. The Bence-Jones proteins do not usually contain carbohydrate. However, if an as-

paragine appears in an Asn-X-Thr sequeon in the V region of the polypeptide chain (see Section 3.3), glycosylation will then occur on this sequeon.

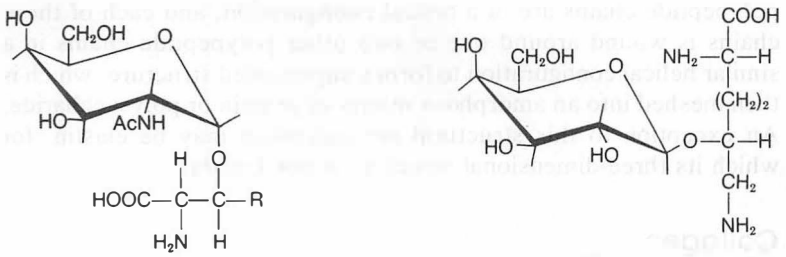
Type II linkages involve glycosidic carbohydrate linkages to serine and threonine alcoholic R groups (Figure 3.20). This type of linkage is found in glycoproteins in mucus secretions, including those from the submaxillary glands, the epithelial cells of the gastrointestinal tract, the respiratory tract, and the female genital tract. No particular sequeon is required for enzymes forming the type II bond. Type II β-D-xylopyranosyl bonds occur in the proteoglycans of connective tissue and in the important protein heparin secreted by mast cells into blood where it acts as an anticoagulant.

Type III linkages made to 5-hydroxylysine (Figure 3.20) are found in some mammalian collagens (see Section 3.6), and in the serum complement protein C1q (see Clin. Corr. 3.1). The 5-hydroxylysine is a derived amino acid (Chapter 2) formed from lysines after incorporation of the lysine into the polypeptide chains.

Other types of covalent sugar : protein linkages are *type IV*, made to 4-hydroxyproline side chains (4-hydroxyproline is also a derived amino acid); *type V*, made to cysteine side chains; and *type VI*, made to NH_2-terminal amino groups. These linkages are not widely dis-

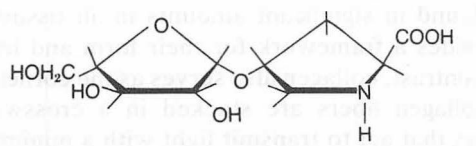


Type I



Type II

Type III

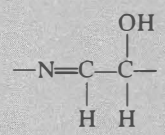


Type IV

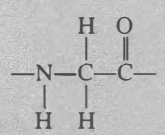
Figure 3.20
Examples of glycosidic linkages to amino acids in proteins.
Type I is an N*-glycosidic linkage through an amide nitrogen of Asn; type II is an* O*-glycosidic linkage through the OH of Ser* ($R = H$) or Thr ($R = CH_3$); *type III is an* O*-glycosidic linkage to the 5-OH of 5-hydroxylysine; type IV is an* O*-glycosidic linkage through 4-OH of 4-hydroxyproline.*
Diagrams modified from R. D. Marshall in *Chemistry of Macromolecules*, IIB, R. E. Offord, ed., University Park Press, Baltimore, Md., 1979, p. 12.

CLIN. CORR. 3.7 GLYCOSYLATED HEMOGLOBIN, HbA_{1c}

A glycosylated hemoglobin, designated Hb_{1c}, is formed spontaneously in the red blood cell by combination of the NH_2-terminal amino groups of the hemoglobin β chain and glucose. The aldehydic function in the glucose first forms a Schiff base with the NH_2-terminal amino group,



which then rearranges to a more stable amino ketone linkage,



by a spontaneous (nonenzymatic) reaction known as the Amadori rearrangement. The concentration of Hb_{1c} is dependent on the concentration of glucose in the blood, and in prolonged hyperglycemia may rise to 12% or more of the total hemoglobin. Patients with diabetes mellitus will tend to have high concentrations of glucose and therefore high amounts of Hb_{1c}. The changes in the concentration of Hb_{1c} in diabetic patients can be used to follow the effectiveness of treatment for the diabetes.

tributed in the glycoproteins of mammals, although type V linkages are found in some glycoproteins obtained from the red blood cell membrane, and type VI are found in glycosylated adult hemoglobin, Hb_{1c} (Clin. Corr. 3.7). Type IV linkages have been found only in higher plants.

3.6 COLLAGEN AND THE FIBROUS PROTEINS

In a previous section hemoglobin and myoglobin were described as examples of globular proteins. Characteristic of the globular proteins are a spheroid shape, varied molecular weights, relatively high water solubility, and a variety of functional roles such as catalysis, transport, and physiological and genetic control. In contrast, characteristics of the fibrous proteins are a rodlike shape, a high amount of secondary structure, and a structural rather than a chemical role in the mammalian organism.

The fibrous proteins form the structural matrix of mammalian bone, ligaments, and skin. These proteins have an unusually high tensile strength. For example, the protein collagen can withstand an amazingly high longitudinal stretching force of up to 1,000,000 psi without tearing. Other fibrous proteins are a part of muscle, where they can interdigitize in a dynamic way that leads to muscle contraction. Another of the fibrous proteins can stretch and reform like rubber, and provides the elasticity of the arteries and the ligaments.

Characteristic of many of the fibrous proteins is that their polypeptide chains are in a helical configuration, and each of these chains is wound around one or two other polypeptide chains in a similar helical configuration to form a supercoiled structure, which is then meshed into an amorphous matrix of protein or polysaccharide. An exception to this structural generalization may be elastin, for which its three-dimensional structure is not known.

Collagen

Distribution of Collagen in Humans

Collagen is found in significant amounts in all tissues and organs, where it provides a framework for their form and lends structural strength. In contrast, collagen also serves as the cornea of the eye, in which the collagen fibers are stacked in a crosswise fashion in stacked arrays that act to transmit light with a minimum of scattering. The percentage of collagen by weight for some representative

Table 3.12 Comparison of the Amino Acid Content of Human Skin Collagen (Type I) and Mature Elastin with That for Two Typical Globular Proteins[a]

Amino Acid	*Collagen (Human Skin)*	*Elastin (Mammalian)*	*Ribonuclease (Bovine)*	*Hemoglobin (Human)*
COMMON AMINO ACIDS		PERCENT OF TOTAL		
Ala	11	22	8	9
Arg	5	0.9	5	3
Asn			8	3
Asp	5	1	15	10
Cys	0	0	0	1
Glu	7	2	12	6
Gln			6	1
Gly	33	31	2	4
His	0.5	0.1	4	9
Ile	1	2	3	0
Leu	2	6	2	14
Lys	3	0.8	11	10
Met	0.6	0.2	4	1
Phe	1	3	4	7
Pro	13	11	4	5
Ser	4	1	11	4
Thr	2	1	9	5
Trp	2	1	9	2
Tyr	0.3	2	8	3
Val	2	12	8	10
DERIVED AMINO ACIDS				
Cystine	0	0	7	0
3-Hydroxyproline	0.1	—	0	0
4-Hydroxyproline	9	1	0	0
5-Hydroxylysine	0.6	0	0	0
Desmosine and isodesmosine	0	1	0	0

[a] Boxed numbers emphasize important differences in amino acid composition between the fibrous proteins (collagen and elastin) and typical globular proteins.

tissues and organs is 4% of liver, 10% of lung, 12 to 24% of the aorta, 50% of cartilage, 23% of whole cortical bone, 68% of the cornea, and 72% of skin. As such, collagen is the most abundant protein in the human organism (Clin. Corr. 3.8).

Collagen is synthesized by a variety of specialized cells in the organs and tissues. These include the fibroblast cells of connective tissue and tendons, osteoblasts in bone, chondroblasts in embryonic

CLIN. CORR. 3.8 SYSTEMIC SYMPTOMS IN DISEASES OF COLLAGEN

As collagen is found in almost all tissues and organs, diseases of collagen biosynthesis or of collagen structure are systemic. Thus diseases resulting from either insufficient synthesis of collagen or the synthesis of structurally malformed collagen show multiorgan symptoms. These symptoms include cardiovascular abnormalities, bone structure (skeletal) abnormalities, skin lesions and other epithelial tissue abnormalities, and eye lens malfunction.

Diseases of collagen biosynthesis or structure include Marfan's syndrome, Ehlers-Danlos syndrome, and scurvy (vitamin C deficiency).

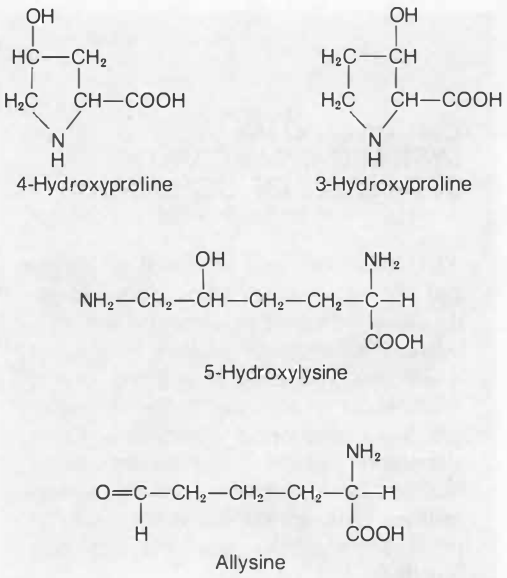


Figure 3.21
Derived amino acids found in collagen. *Carbohydrate is attached to 5-OH in 5-hydroxylysine by a type III glycosidic linkage (see Figure 3.20).*

cartilage, and odontoblasts in teeth. The biosynthesis of collagen is described in Chapter 19.

The collagen fibrils in the tissues are usually imbedded in an amine-containing polysaccharide or a glycoprotein matrix with a gellike character, known as the *ground substance.*

Amino Acid Composition of Collagen

The amino acid composition of collagen is very different from that found for a typical globular protein (Table 3.12). Its amino acid composition is high in glycine (~33% of the total), proline (~10%), the derived amino acid hydroxyproline (~10%), and the derived amino acid 5-hydroxylysine (~1%). The predominant form of hydroxyproline (Hyp) is the 4-OH derivative (Figure 3.21), although some types of collagen contain small amounts of the 3-OH form of Hyp. A small amount of carbohydrate is found in collagen covalently bonded through the 5-OH of 5-hydroxylysines. Formation of the derived amino acids and addition of the carbohydrate occurs intracellularly by an enzyme-catalyzed process prior to secretion of the collagen from its synthesizing cell (see Chapter 19).

Amino Acid Sequence of Collagen

The molecular unit of collagen in collagen fibrils is the *tropocollagen* molecule. Each tropocollagen molecule contains three polypeptide chains. In the tropocollagen of some tissue types, each of the three polypeptide chains has an identical amino acid sequence (e.g., type II collagen found in cartilage; type III collagen found in cartilage, scar, and soft tissue; type IV collagen found in basement membrane). In the tropocollagen classified as type I, two of the chains are identical in sequence and the amino acid sequence of the third chain differs slightly (e.g., type I collagen found in bone, tendons, soft tissue, and scar tissue). For the latter collagen, the two identical chains are designated α_1 chains, and the third nonidentical chain, α_2. The different types of collagen (types I to IV) are characterized by differences in their physical properties due to the distinct differences in their amino acid sequence and in their percent carbohydrate (Table 3.13).

The amino acid sequence of the polypeptide strands in the tropocollagen molecule is also quite different from the sequences of the globular proteins. It is found that, except for a small segment of 15 to 25 amino acids on the NH_2- and COOH-terminal ends (these segments are known as the telopeptides), *glycine is found every third amino acid in the sequence.* Furthermore, *the sequences Gly-Pro-Y and Gly-X-Hyp* (where X and Y are any of the amino acids) *are each repeated about 100 times* in the polypeptide sequence of a chain. These two triplet sequences will thus encompass about 600 amino

acids within a tropocollagen chain of approximately 1,000 amino acids.

Structure of Tropocollagen

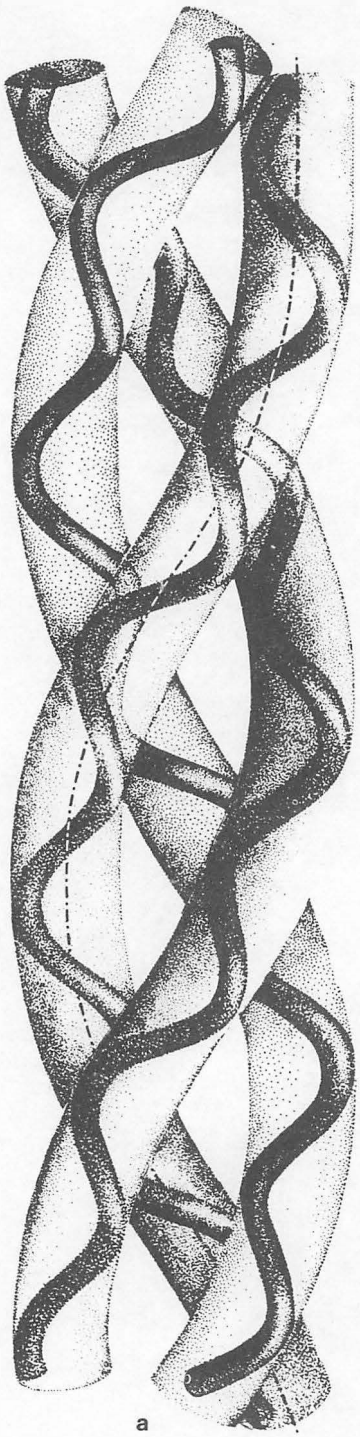
Polypeptides that contain only the amino acid proline (polyproline polymer) can be synthesized in the laboratory. This polyproline polypeptide will spontaneously assume a regular secondary structure in aqueous solution in which the polypeptide chain is in a tightly twisted extended helix with three residues per turn of the helix ($n = 3$), $d = 2.9$ Å, and $p = 8.7$ Å. This helix with *trans* peptide bonds is designated the *polyproline type II helix*. The helix conformation found for polyproline has the same characteristics as the helix found for each of the polypeptide chains in the native tropocollagen, in which the primary structure contains a proline or hydroxyproline at approximately every third position within the amino acid sequence.

Since the same helix is found in the polypeptide chains in collagen as in polyproline, the thermodynamic forces leading to the formation of this helix must be based on the properties of proline. In proline the ϕ angle is part of a five-membered ring, which constrains it to an angle compatible with the collagen helix. In addition, the hydrodynamic solvation properties of Pro and Hyp stabilize the collagen helix conformation.

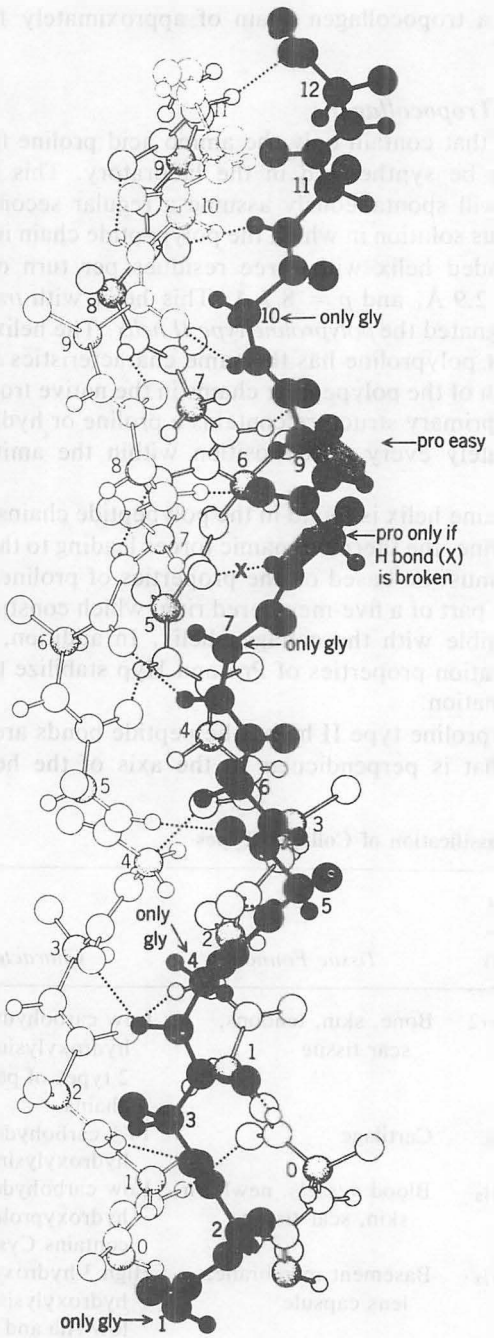
In the polyproline type II helix, the peptide bonds are positioned in a plane that is perpendicular to the axis of the helix. In this

Table 3.13 Classification of Collagen Types

Type	*Chain Designations*	*Tissue Found*	*Characteristics*
I	$[\alpha 1(I)]_2\alpha 2$	Bone, skin, tendons, scar tissue	Low carbohydrate; <10 hydroxylysines per chain; 2 types of polypeptide chains
II	$[\alpha 1(II)]_3$	Cartilage	10% carbohydrate; >20 hydroxylysines per chain
III	$[\alpha 1(III)]_3$	Blood vessels, newborn skin, scar tissue	Low carbohydrate; high hydroxyproline and Gly; contains Cys
IV	$[\alpha 1(IV)]_3$	Basement membrane, lens capsule	High 3-hydroxyproline; >40 hydroxylysines per chain; low Ala and Arg; contains Cys



a


12
11
10
9
8
7
6
5
4
3
2
1
0
10 ← only gly
← pro easy
9
8
6
7
pro only if
H bond (X)
is broken
X
5
6
7
← only gly
4
5
6
3
4
5
only
gly
2
3
4
1
2
3
1
0
2
0
only gly → 1
← 5 Ångstroms →

(b)

geometry the peptide carbonyl groups are pointed in a direction where they can form strong *interchain* hydrogen bonds to a parallel polypeptide chain. This is in contrast to the orientation of the peptide bond in the α helix configuration, in which the peptide bond is directed in a plane parallel to the axis of the α helix and form *intrachain* hydrogen bonds with other peptide bonds in the same polypeptide chain.

Three polypeptide chains, each in a polyproline type II helix conformation, are wound about each other in a defined way (Figure 3.22) to form a superhelix structure of the tropocollagen molecule. This superhelix forms because glycine residues, with a small side chain group (R = H), appear at every third position in the amino acid sequence of each of the polypeptide chains. Since the helix has three amino acids per turn of the helix ($n = 3$), the glycines are positioned along the same side of the helix and thus form an apolar longitudinal edge along the outside of the helix of the polypeptide chain. This allows each of the three strands to come close to one another (because the glycine side chain is small) in a regular pattern, with the supercoil structure stabilized by the hydrophobic interstrand interactions between the apolar edges (Figure 3.22). The interchain hydrogen bonds between peptide groups further stabilize the supercoil structure. A larger side chain group than glycine would sterically prevent the adjacent strands from coming together in the superhelix configuration.

The three polypeptide chains that form a supercoiled structure are initially tied together by covalent cystine bonds formed between chains in the COOH-terminal region of procollagen (biosynthetic precursor form of collagen; see Chapter 19). This region is hydrolyzed from the polypeptide chains during the final step in the synthesis of tropocollagen. The supercoiled tropocollagen molecule is approximately 15 by 3,000 Å and has a molecular weight of 300,000.

←

Figure 3.22
Diagram of tropocollagen demonstrating the necessity for glycine every third residue to allow the different chains to be in close proximity in the structure. All α carbons are numbered and proposed hydrogen bonds are shown by dotted lines.
(a) Ribbon model for supercoiled structure of tropocollagen with each of the individual chains in a polyproline type II helix. (b) More detailed model of supercoiled configuration.
Reprinted with permission from R. E. Dickerson and I. Geis, *The Structure and Actions of Proteins*, W. A. Benjamin, Inc., Menlo Park, 1969, pp. 41, 42.

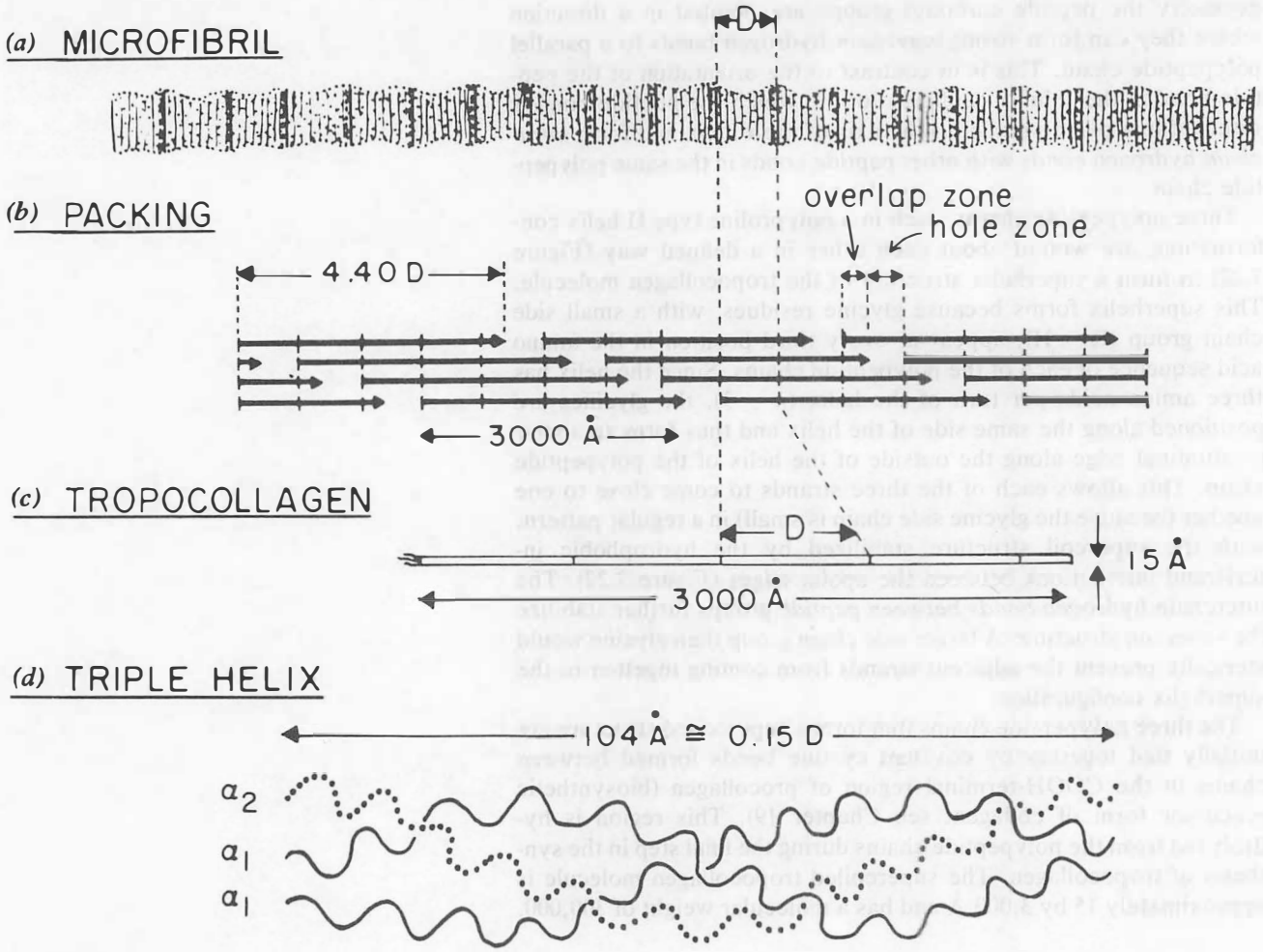


Figure 3.23
D-Spacing of collagen.
(a) Segments are stained with positive electron-dense stains which visualizes D-periodicity. (b) Packing of tropocollagen molecules in microfibril with quarter staggered arrangement and hole for hydroxyapatite initiation. (c) Tropocollagen molecules dimensions. (d) Helix of tropocollagen showing three chains for type I collagen.

Reprinted with permission from M. E. Grant and D. S. Prockop, *New Engl. J. Med.* **286,** 194 (1972).

Assembly of the Collagen Microfibril

Characteristic of the collagen are regions that appear every 670 Å along the axis of the supercoiled tropocollagen molecule of amino acids containing polar R groups. Since these regions have a periodicity of 670 Å in the 3,000 Å long molecule, the regions of polar residues will appear four times along the axis of the molecule (Figure 3.23). The length of 670 Å is known as the collagen D period.

Tropocollagen molecules assemble into a fibril in a specific arrangement that is stabilized by the interactions between D period-spaced polar and intervening nonpolar regions along the tropocollagen axis. In this assembly each tropocollagen molecule associates in the fibril with its NH_2-terminal end one D period down from the

NH_2-terminal end of the adjacent tropocollagen molecule (Figure 3.23). This gives a quarter staggered arrangement of the tropocollagen molecules in the fibril array, with the polar regions aligned in register (Figure 3.23).

With the D period of 670 Å in a molecule 3,000 Å long, there exists a nonintegral number of D period lengths of 4.4 to 4.5 D periods per tropocollagen unit. In order to maintain a configuration in register, there is a gap in a row of the fibril assembly, between the COOH-terminal end of one tropocollagen molecule and the next NH_2-terminal end of the next tropocollagen molecule. This gap is 300 to 370 Å (one-half a D period) (Figure 3.23). It is thought that these gaps serve as the nucleation sites for the initiation of hydroxyapatite $[Ca_2(PO_4)_3]$ crystallization in bone formation.

Formation of Covalent Cross-Links in Collagen

An enzyme present in the collagen fiber (see Chapter 19) acts on some of the lysines in tropocollagen to catalyze the oxidative deamination of the ε-amino groups to a δ-aldehyde (Figure 3.24). The resulting derived amino acid, containing an aldehydic side chain group R, is known as *allysine*. This aldehyde function will spontaneously undergo nucleophilic addition reactions with the ε-amino groups of other (nontransformed) lysines, and the δ-carbon of other allysine side chains to form covalent interchain bonds (Figure 3.24). These

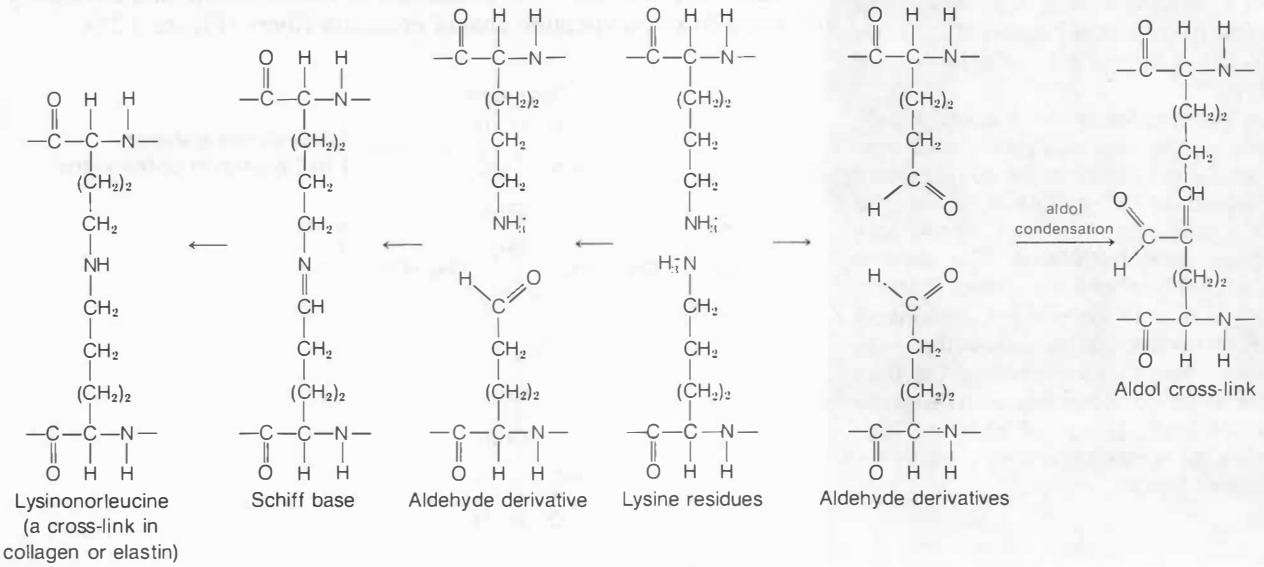


Figure 3.24
Covalent cross-links formed in collagen from allysine.

CLIN CORR. 3.9 DISEASES RELATED TO THE INABILITY TO FORM COVALENT CROSS-LINKS IN COLLAGEN

The cross-linking of tropocollagen molecules in collagen fibers is prevented in certain diseases.

Lathyrism is a disease caused by the ingestion of a poison such as α-aminopropionitrile ($H_2N—CH_2—CH_2—CN$) found in the *Lathyrus odoratus* pea plant. As covalent cross-links in collagen will not occur on the inhibition of lysyl oxidase the collagen formed is unstable and turns over rapidly. Symptoms of lathyrism include skeletal deformities and a high excretion of hydroxyproline in urine due to the increased rates of collagen turnover.

In Ehlers-Danlos syndrome (type V), there is thought to be a deficiency in the enzyme lysine amino oxidase. The symptoms of this disease are similar to those of lathyrism.

In homocystinuria there exists a deficiency of the enzyme cystathionine synthetase, which catalyzes the conversion of homocysteine to cystathionine in the biosynthetic pathway that synthesizes cysteine from methionine. The derived amino acid homocysteine, which is structurally similar to cysteine but contains an extra methylene carbon beween the —SH and α-carbon, thus accumulates. The thiol sulfur of the homocysteine reacts with the free aldehydic groups of allysines, preventing the normal cross-linking within the collagen fibers.

covalent bonds may be to adjacent tropocollagen molecules as well as between strands within the same tropocollagen molecule. The covalent cross-links formed with allysines are essential to the stability of the fibril structure. If the covalent cross-links are not formed, the tropocollagen fibrils are easily dissociated and degraded. (Clin. Corr. 3.9).

Tropoelastin

The protein elastin gives tissues and organs the capacity to stretch without tearing. Elastin is found in high concentrations in the ligaments, lungs, walls of the arteries, and skin.

The amino acid composition of elastin shows it to be approximately one-third glycine and one-third proline. However, in contrast to collagen, only very small amounts of hydroxyproline are found. Elastin strands do not form a polyproline-type helix. Unfortunately, the actual conformation of elastin is still unknown.

An extracellular enzyme acts in elastin on particular lysine R groups, in a similar way as in collagen, to convert these lysines to allysines. In particular, the allysine residues appear to be generated where lysines appear in the sequence *Lys*-Ala-Ala-*Lys* and *Lys*-Ala-Ala-Ala-*Lys* in the primary structure. When the pair of lysines or allysines in this sequence from adjacent polypeptide chains comes close, three allysines and one lysine residue can react to form the heterocyclic structure of desmosine or isodesmosine that covalently cross-links polypeptide chains in elastin fibers (Figure 3.25).

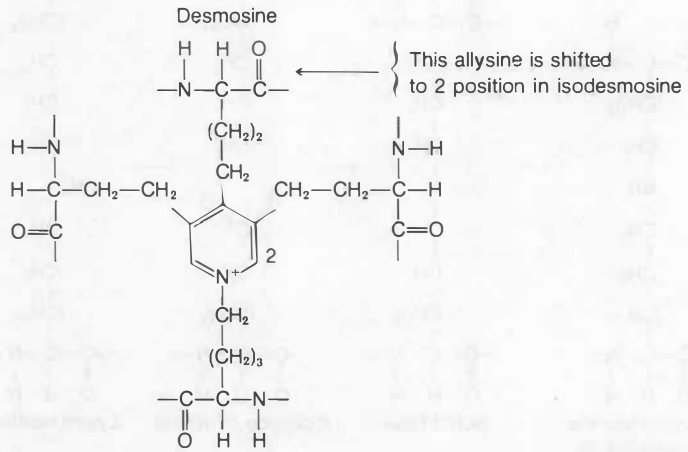


Figure 3.25
Desmosine and isodesmosine covalent cross-link in elastin.

α-Keratin

α-Keratin is found in the epidermal layer of the skin, in nails, and in the hair. The conformation of each of the polypeptide chains in α-keratin is α-helical.

The analysis of the primary structure of the α-keratin polypeptide chain shows a section of repetitive seven amino acid segments, of which the first and fourth amino acids in the seven-member segment are hydrophobic; and the fifth and seventh members are polar and often of opposite sign. The seven-amino acid repetitive sequence can be represented by the formula (a-b-c-d-e-f-g), where a and d contain hydrophobic side chain groups, and e and g contain polar or ionized side chain groups. Since a seven-amino acid segment represents approximately two complete 360° turns about the α helix ($n = 3.6$), the apolar residues a and d will align in the helix to form an apolar edge along one side of the helix (Figure 3.26), similar to the way in which every third residue of glycine in a polyproline type II helix forms an apolar edge on its helix. The apolar edge interacts with similar edges on other α-helical polypeptide chains of α-keratin proteins to form a superhelical structure of two or three polypeptide chains. The polar residues in each seven-membered repetitive sequence, which form polar edges to the individual strands, act to stabilize the multistrand supercoiled structure. The supercoil helix of α-keratin is also stabilized by covalent peptide cross-links formed between glutamine side chains and lysine ε-NH_2 groups (Figure 3.27).

The supercoiled molecular structures (protofibrils) are a part of a larger assembly in which the protofibrils are wound around one another to produce an even larger structure, designated a microfibril. The microfibril is 70 Å in diameter and may be composed of nine protofibrils (Figure 3.28).

The α-keratin protein contains a high concentration of cysteine that acts to form disulfide cross-links between polypeptide strands within the fibril structure and an amorphous protein (keratohyalin) that forms a matrix of high cysteine content in which the α-keratin microfibrils are imbedded. An analogy may be made to reinforced concrete, with the α-keratin serving as the metal rods and the keratohyalin protein as the cement. The cysteine content is higher in the keratin of nails and hair than in the keratin of epidermis tissue.

Tropomyosin

Tropomyosin is a component of the thin filament of muscle tissue (Figure 3.29). Like α-keratin, each polypeptide chain in the tropomyosin molecule is in an α-helical configuration. Also as in α-keratin, an analysis of its primary structure shows a type of repeat-

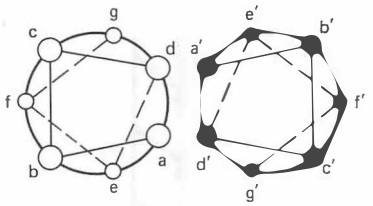


Figure 3.26
Interaction of a polar edge of two chains in α-helical configuration as in α-keratin and in tropomyosin.
Interaction of apolar a–d' and d–a' residues of two α helices aligned parallel in an NH_2-terminal (top) to COOH-terminal direction is presented.
From A. D. McLachlan and M. Stewart, *J. Mol. Biol.*, **98**, 293 (1975).

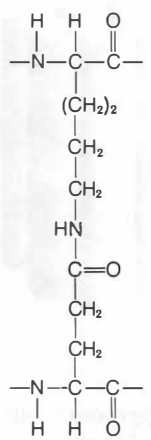


Figure 3.27
Amide cross-link formed between Gln and Lys in α-keratin.

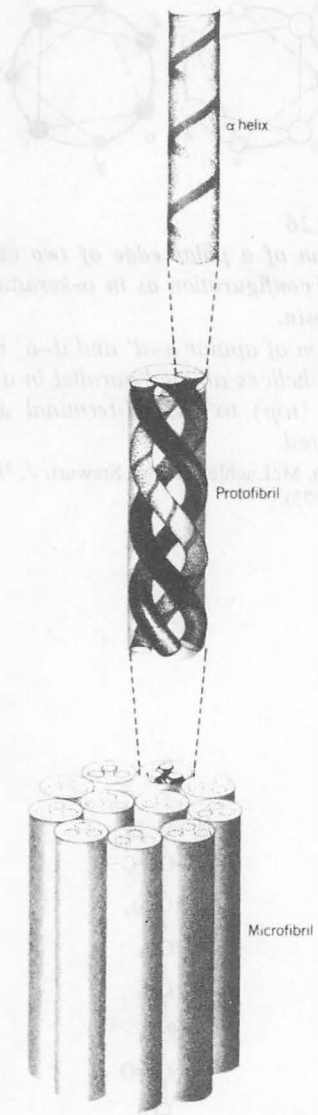


Figure 3.28
Assembly of polypeptides into supercoils and microfibrils of α-keratin.
Reprinted with permission from R. E. Dickerson and I. Geis, *The Structure and Action of Proteins*, W. A. Benjamin, Inc., Menlo Park, 1969, p. 37.

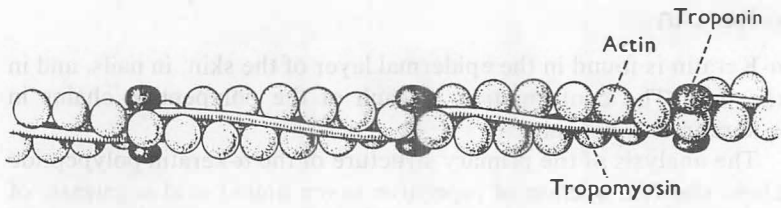


Figure. 3.29
Structure of the thin filament of muscle tissue with tropomyosin strands (containing two polypeptide chains) showing association of actin monomers and troponin.
Reprinted with permission from S. Ebashi, M. Endo, and I. Ohtsuki, *Q. Rev. Biophys.*, **2**, 351 (1969).

ing seven-membered amino acid sequence in terms of apolar and polar side chain R groups $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$ with residues a and d apolar and residues e and g ionized. Two strands with this α-helical sequence are wound around each other into a left-handed superhelical coil.

The tropomyosin molecule is 400 Å long, with the 400 Å length divided into approximately 14 regions of 28 Å. Within each of these regions there is a zone with a high concentration of positively charged side chain groups, a zone with a high concentration of negatively charged side chain groups, and a zone with a high concentration of apolar side chain groups. These zones are thought to provide binding sites for the other protein components of muscle (actin and troponin) that associate with tropomyosin in a regular pattern (Figure 3.29).

Summary of Fibrous Protein Structures

In collagen, α-keratin, and tropomyosin, multiple polypeptide chains with a highly regular secondary structure, are found wound around each other in a supercoil configuration. In turn, the coils are wound or aligned in fibers that are stabilized by covalent cross-links in collagen and α-keratin. The amino acid sequence of the polypeptide chains are repetitive, generating edges on the cylindrical surface of these helical-shaped molecules that stabilize hydrophobic interactions between strands in a supercoiled array. In addition, collagen and α-keratin are immersed in a matrix of polysaccharide or high cysteine-containing protein that cements these fibers into the matrix of the animal tissue or organ to which they give structure and strength.

BIBLIOGRAPHY

Hemoglobin

Baldwin, J., and Chothia, C. Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. *J. Mol. Biol.* 129:175, 1979.

Perutz, M. Structure and mechanism of hemoglobin. *Br. Med. Bull.* 32:195, 1976.

Perutz, M. Hemoglobin structure and respiratory transport. *Sci. Am.* 239(6):92, 1978.

Antibodies

Kabat, E. A. The structural basis antibody complementarity. *Adv. Protein Chem.* 32:1, 1978.

Poljak, R. J. Correlations between three dimensional structure and function of immunoglobins. *CRC Crit. Rev. Biochem.* 5:45, 1978.

Glycoproteins

Clamp, J. R. Structure and function of glycoproteins, in F. W. Putnam (ed.), *The plasma proteins,* vol. II. New York: Academic, 1975, p. 163.

Marshall, R. D. Structures and functions of glycoproteins in R. E. Offord (ed.), *Chemistry of macromolecules IIB.* Baltimore: University Park Press, 1979, p. 1.

Marshall, R. D. Some observations on why many proteins are glycosylated. *Biochem. Soc. Trans.* 7:800, 1979.

Plasma Lipoproteins

Osborne, J. C., Jr., and Brewer, H. B., Jr. The plasma lipoproteins. *Adv. Protein Chem.* 31:253, 1977.

Scanu, A. M. Plasma lipoprotein: structure, function and regulation. *Trends Biochem. Sci.* 3:202, 1978.

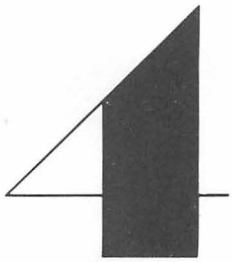
Souter, A. K., and Myant, N. B. Plasma lipoproteins, in R. E. Offord (ed.), *Chemistry of macromolecules IIB.* Baltimore: University Park Press, 1979, p. 55.

Fibrous Proteins

Bornstein, P., and Traub, W. The chemistry and biology of collagen, in H. Neurath and R. L. Hill (eds.), *The proteins,* 3rd ed., vol. IV. New York: Academic, 1979, p. 411.

Miller, A. Structure and function of fibrous proteins, in R. E. Offord (ed.), *Chemistry of macromolecules IIA*. Baltimore: University Park Press, 1979, p. 171.

Smillie, L. B. Structure and function of tropomyosins from muscle and non-muscle sources. *Trends Biochem. Sci.* 4:151, 1979.



Enzymes

J. LYNDAL YORK

CLASSIFICATION, KINETICS, AND CONTROL

4.1 GENERAL CONCEPTS

Enzymes are proteins evolved by the cells of living organisms for the specific function of catalyzing chemical reactions. Enzymes increase the rate at which reactions approach equilibrium. *Rate* is defined as the change in the amount (moles, grams) of starting materials or products per unit time. The enzyme triggers the increased rate by acting as a catalyst. A true catalyst increases the rate of a chemical reaction, but is not itself changed in the process. The enzyme may become temporarily covalently bound to the molecule being transformed during intermediate stages of the reaction, but in the end the enzyme will be regenerated in its original form as the product is released.

There are two important characteristics of catalysts in general and enzymes in particular that should not be forgotten. The first is that the enzyme is not changed by entering into the reaction. The second is that the enzyme does not change the equilibrium constant of the reaction, it simply increases the rate at which the reaction approaches equilibrium. Therefore, a catalyst is responsible for increasing the rate but not changing the thermodynamic properties of the system with which it is interacting. In biological systems, a catalyst is necessary because at the temperature and pH of the human body, reactions would not occur at a rate sufficient to support rapid muscular activity, nerve impulse generation, and all the other processes required to support life.

At this point, we need to define several terms before entering into a discussion of the mechanism of enzyme action. An *apoenzyme* is the protein part of the enzyme minus any cofactors or prosthetic groups that may be required for the enzyme to be functionally active. The apoenzyme is therefore catalytically inactive. Not all enzymes require cofactors or prosthetic groups to be active. The *cofactors* are those small organic or inorganic molecules that the enzyme requires for its activity. For example, lysine oxidase is a copper-requiring enzyme. Copper in this case is loosely bound but is required for the enzyme to be active. The *prosthetic group* is similar to the cofactor but is tightly bound to the apoenzyme. For example, in the cytochromes, the heme prosthetic group is very tightly bound and requires strong acids to disassociate it from the cytochrome. The addition of cofactor or prosthetic group to the apoprotein yields the *holoenzyme,* which is the active enzyme. The molecule the enzyme acts upon to form product is called the *substrate.* Since most reactions are reversible, the products of the forward reaction will become the substrates of the reverse reaction. Enzymes have a great deal of specificity. For example, glucose oxidase will oxidize glucose

but not galactose. This specificity resides in a particular region on the enzyme surface called the *substrate binding site,* which is a particular arrangement of chemical groups on the enzyme surface that is specially formulated to bind a specific substrate. The substrate binding site may have integrated within it the *active site.* In some cases the active site may not be within the substrate binding site but may be contiguous to it in the primary sequence. In other cases the active site lies in distant regions of the primary sequence but is brought adjacent to the substrate binding site by folding of the tertiary structure. The active site contains the machinery, in the form of particular chemical groups, that is involved in catalyzing the reaction under consideration. The chemical groups involved in both binding of substrate and catalysis are often part of the side chains of the amino acids of the apoenzyme.

In some enzymes there is another region of the molecule, the *allosteric site,* that is not at the active site or substrate binding site, but is somewhere else on the molecule. The allosteric site is the site where small molecules bind and effect a change in the active site or the substrate binding site. The binding of a specific small organic molecule at the allosteric site causes a change in the conformation of the enzyme, and that conformational change may cause the active site to become either more active or less active. It may cause the binding site to have a greater affinity for substrate, or it may actually cause the binding site to have less affinity for substrate. Such interactions are involved in the regulation of the activity of enzymes and is discussed in more detail on page 200.

4.2 CLASSIFICATION OF ENZYMES

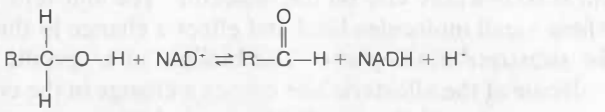
The International Union of Biochemistry has established a system whereby all enzymes are placed into one of six major classes. Each class is then subdivided into several subclasses, which are further subdivided. A number is assigned to each class, subclass, and subsubclass so that an enzyme is assigned a four-digit number as well as a name. The fourth digit identifies a specific enzyme. For example, alcohol: NAD oxidoreductase is assigned the number 1.1.1.1. because it is an oxidoreductase, the electron donor is an alcohol and the acceptor is the coenzyme NAD. Notice that in naming an enzyme, the substrates are stated first, followed by the reaction type to which the ending *ase* is affixed. The trivial name of the enzyme 1.1.1.1. is alcohol dehydrogenase. Many common names persist but

are not very informative. For example, "aldolase" does not tell much about the substrates, although it does identify the reaction type. We will use the trivial names which are recognized by the I.U.B.

Each of the six major enzyme classes will be briefly described in the following sections.

Class 1. Oxidoreductases

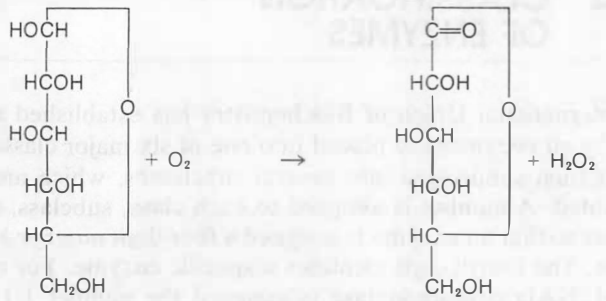
These enzymes are involved in oxidation and reduction. For example, alcohol : NAD oxidoreductase catalyzes the oxidation of an alcohol to an aldehyde. This enzyme removes two electrons and two hydrogens from the alcohol to yield an aldehyde, and in the process, the two electrons that were originally in the carbon–hydrogen bond of the alcohol are transferred to the NAD, which then becomes reduced.



In addition to the alcohol and aldehyde functional groups, the *dehydrogenases* also act on the following functional groups as electron donors: $-CH_2-CH_2-$, $-CH_2-NH_2$, $-CH=NH$, as well as the nucleotides NADH and NADPH.

Other major subclasses of the oxidoreductases are summarized as follows:

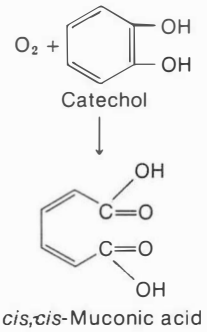
Oxidases transfer two electrons from the donor to oxygen, resulting in hydrogen peroxide formation. For example, glucose oxidase catalyzes the reaction shown here.



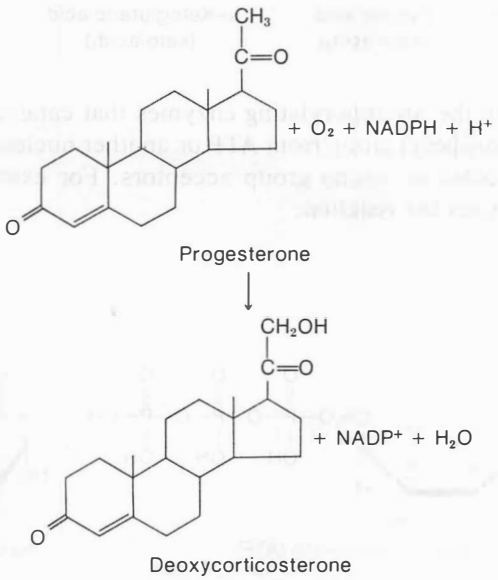
β-D-Glucose δ-Gluconolactone

In the case of the cytochrome oxidases, H_2O rather than H_2O_2 is the product. These are specialized oxidases that transfer electrons from reduced nucleotides (NADH) to oxygen.

Oxygenases catalyze the incorporation of both atoms of oxygen into a single substrate. Catechol oxygenase catalyzes the reaction shown to the right.



Hydroxylases incorporate one atom of molecular oxygen into the substrate; the second oxygen appears as water. The steroid hydroxylases typify this reaction type.



Peroxidases utilize hydrogen peroxide rather than oxygen as the oxidant. NADH peroxidase catalyzes the reaction

$$NADH + H^+ + H_2O_2 \rightleftharpoons NAD^+ + 2H_2O$$

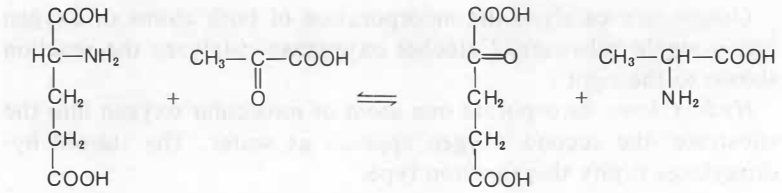
Catalase is unique in that hydrogen peroxide serves as both donor and acceptor. Catalase functions in the cell to detoxify hydrogen peroxide:

$$H_2O_2 + H_2O_2 \rightleftharpoons O_2 + 2H_2O$$

Class 2. Transferases

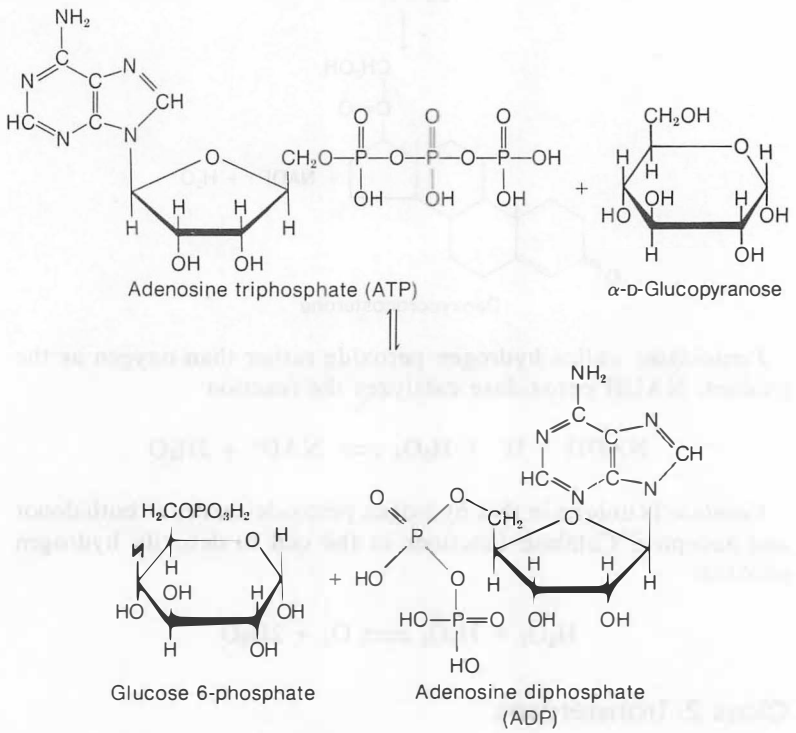
These enzymes are involved in transferring functional groups between donors and acceptors. The amino, acyl, phosphate, one-carbon, and glycosyl groups are the major moieties that are transferred.

Transaminases transfer the amino group from one amino acid to a keto acid acceptor, resulting in the formation of a new amino acid.



L-Glutamic acid (amino acid$_1$) | Pyruvic acid (keto acid$_2$) | α-Ketoglutaric acid (keto acid$_1$) | L-Alanine (amino acid$_2$)

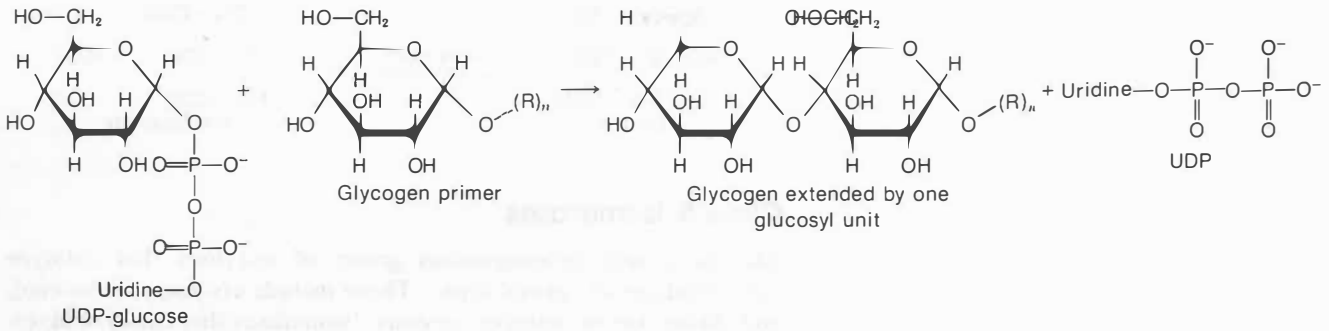
Kinases are the phosphorylating enzymes that catalyze the transfer of the phosphoryl group from ATP or another nucleotide triphosphate, to alcohol or amino group acceptors. For example, glucokinase catalyzes the reaction:



Adenosine triphosphate (ATP) | α-D-Glucopyranose

Glucose 6-phosphate | Adenosine diphosphate (ADP)

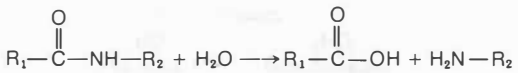
The synthesis of glycogen depends upon *glycosyltransferases*, which catalyze the transfer of an activated glucosyl residue to a glycogen primer. The phosphoester bond in uridine diphosphoglucose is labile, which allows the glucose to be transferred to the growing end of the glycogen primer, that is,



It should be noted that although a polymer is synthesized, the reaction is not of the ligase type, which we discuss in Class 6 below, as is the formation of protein from activated amino acids.

Class 3. Hydrolases

This group of enzymes can be considered as a special class of the transferases in which the donor group is transferred to water. The generalized reaction involves the hydrolytic cleavage of C—O, C—N, O—P, and C—S bonds. The cleavage of the peptide bond is a good example of this reaction:

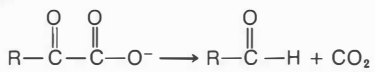


The proteolytic enzymes are a special class of hydrolases called *peptidases.*

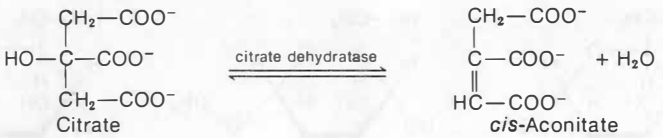
Class 4. Lyases

Lyases are enzymes which add or remove the elements of water, ammonia, or CO_2 to double bonds.

The *decarboxylases* remove the element of CO_2 from α- and β-keto acids or amino acids:

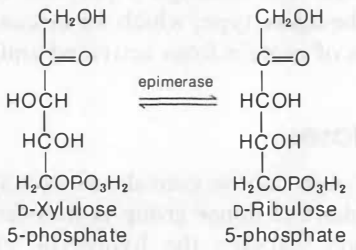


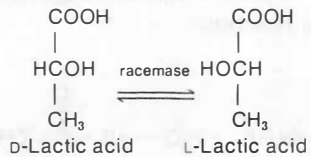
The *dehydratases* remove the elements of H_2O in a dehydration reaction. Citrate dehydratase converts citrate to *cis*-aconitate.



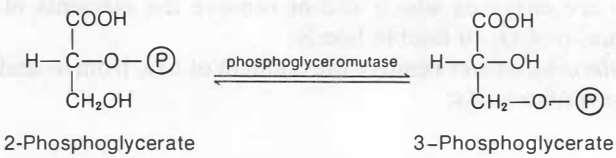
Class 5. Isomerases

This is a very heterogeneous group of enzymes that catalyze isomerizations of several types. These include cis–trans, keto–enol, and aldose–ketose interconversions. Isomerases that catalyze inversion at asymmetric carbons are either *epimerases* or *racemases*.





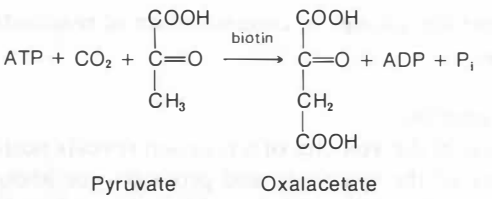
Mutases involve the intramolecular transfer of a group such as the phosphoryl. The transfer need not be direct but can involve a phosphorylated enzyme as an intermediate. An example is phosphoglycerate mutase, which catalyzes the conversion of 2-phosphoglycerate to 3-phosphoglycerate:



Class 6. Ligases

Since ligate means to bind, these enzymes are involved in synthetic reactions where two molecules are joined at the expense of an ATP

"high-energy phosphate bond." The use of "synthetase" is reserved for this particular group of enzymes. The formation of amino acid tRNAs, acyl coenzyme A, glutamine, and the addition of CO_2 to pyruvate, are reactions catalyzed by ligases. Pyruvate carboxylase is a good example of a ligase enzyme. It catalyzes the reaction:



The two substrates carbon dioxide and pyruvate are ligated to form a four-carbon keto acid.

The six enzyme classes and most of the important subclass members are compiled in Table 4.1. The accepted trivial names are used for members of the subclass.

Table 4.1 Summary of the Enzyme Classes and Major Subclasses

1. Oxidoreductases	2. Transferases
Dehydrogenases	Transaldolase and ketolase
Oxidases	Acyl-, methyl-, glucosyl-, and phosphoryltransferase
Reductases	Kinases
Peroxidases	Phosphomutases
Catalase	
Oxygenases	
Hydroxylases	
3. Hydrolases	4. Lyases
Esterases	Decarboxylases
Glycosidases	Aldolases
Peptidases	Ketolases
Phosphatases	Hydratases
Thiolases	Dehydratases
Phospholipases	Synthases
Amidases	Lyases
Deaminases	
5. Isomerases	6. Ligases
Racemases	Synthetases
Epimerases	Carboxylases
Isomerases	
Mutases (not all)	

4.3 KINETICS

Basic Chemical Kinetics

Since enzymes affect the rate of chemical reactions, it is important to understand basic chemical kinetics and how kinetic principles apply to enzyme-catalyzed reactions. *Kinetics* is a study of the rate of change of the initial state of reactants and products to the final state of reactants and products. The term *velocity* is often used rather than rate. Velocity is expressed in terms of change in the concentration of substrate or product per unit time, whereas rate refers to changes in total quantity (moles, grams) per unit time. Biochemists tend to use the two terms interchangeably.

The velocity of a reaction A → P is determined from a progress curve or velocity profile of a reaction. The progress curve can be determined by following the disappearance of reactants or the appearance of product at several different times. Such a curve is shown in Figure 4.1, where product appearance is plotted against time. The slope of tangents to the progress curve yields the instantaneous velocity at that point in time. The initial velocity represents an important parameter in the assay of enzyme concentration, as we learn later. Notice that the velocity constantly changes as the reaction

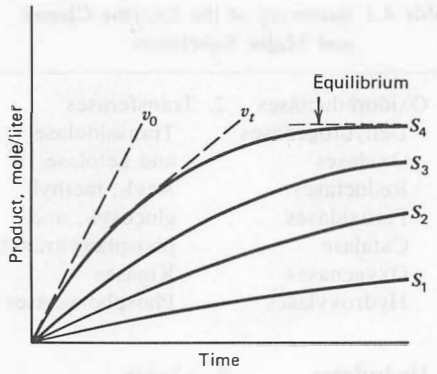


Figure 4.1
Progress curves for an enzyme-catalyzed reaction.
The initial velocity (v_0) *of the reaction is determined from the slope of the progress curve at the beginning of the reaction. The initial velocity increases with increasing substrate concentration* (S_1 *through* S_4) *but reaches a limiting value that is characteristic of each enzyme. The velocity at any time,* t, *is denoted as* v_t.

proceeds to equilibrium, and becomes zero at equilibrium. Mathematically, the velocity is expressed as

$$\text{Velocity} = v = \frac{-d[\text{A}]}{dt} = \frac{d[\text{P}]}{dt} = \frac{-\Delta[\text{A}]}{\Delta t} = \frac{\Delta[\text{P}]}{\Delta t} \quad (4.1)$$

and represents the change in concentration of reactants or products per unit time.

The Rate Equation

Determination of the velocity of a reaction reveals nothing about the stoichiometry of the reactants and products nor about the mechanism of the reaction. What is needed is an equation that relates the experimentally determined initial velocity to the concentration of reactants. Such a relation is the velocity or rate equation. In the case of the reaction A → P, the velocity equation is

$$\frac{-d[\text{A}]}{dt} = v = k[\text{A}]^n \quad (4.2)$$

That is to say, the observed initial velocity will depend on the starting concentration of A to the nth power multiplied by a proportionality constant k. The latter is known as the *rate constant*. The exponent n is usually an integer from 1 to 3 that is required to satisfy the mathematical identity of the velocity expression.

Characterization of Reactions Based on Order

Another term that is useful in describing a reaction is the *order* of reaction. Empirically the order is determined as the sum of the exponents on each concentration term in the rate expression. In the case under discussion the reaction is *first order*, since the velocity depends on the concentration of A to the first power, $v = k[\text{A}]^1$. In a reaction such as A + B → C, if the order with respect to A and B is 1, that is, $v = k[\text{A}]^1[\text{B}]^1$, overall the reaction is second order. It should be noted that the order of reaction is independent of the stoichiometry of the reaction, that is, if the reaction were third order, the rate expression could be either $v = k[\text{A}][\text{B}]^2$ or $v = k[\text{A}]^2[\text{B}]$, depending on the order in A and B. Since the velocity of the reaction is constantly changing as the reactant concentration changes, it is obvious that first-order reaction conditions would not be ideal for assaying an enzyme-catalyzed reaction because one would have two variables, the changing substrate concentration and the unknown enzyme concentration.

If the differential first-order rate expression (4.2) is integrated, one obtains

$$k_1 \cdot t = \frac{2.3 \log [A]}{[A] - [P]} \tag{4.3}$$

where [A] is the initial reactant concentration and [P] is the concentration of product formed at time t. The first-order rate constant k_1 has the units of reciprocal time. If the data shown in Figure 4.1 were replotted as log [P] vs time, a straight line would be obtained whose slope is equal to $k_1/2.303$. The rate constant k_1 should not be confused with the rate or velocity of the reaction.

Many biological processes proceed under first-order conditions. The clearance of many drugs from the blood by peripheral tissues is a first-order process. A specialized form of the rate equation can be used in these cases. If we define $t_{1/2}$ as the time required for the concentration of the reactants or the blood level of a drug to be reduced by one-half the initial value, then equation (4.3) reduces to

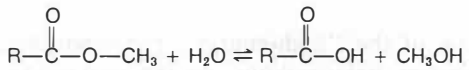
$$k_1 \cdot t_{1/2} = 2.3 \log \left[\frac{1}{1 - \frac{1}{2}} \right] = 2.3 \log 2 = 0.69 \tag{4.4}$$

or

$$t_{1/2} = \frac{0.69}{k_1} \tag{4.5}$$

Notice that $t_{1/2}$ is not one-half the time required for the reaction to be completed. The term $t_{1/2}$ is referred to as the *half-life* of the reaction.

Many *second-order* reactions that involve water or any one of the reactants in large excess can be treated as a pseudo-first-order reaction. In the case of the hydrolysis of an ester,



the second-order rate expression is

$$\text{velocity} = v = k_2[\text{ester}]^1[H_2O]^1 \tag{4.6}$$

but since water is in abundance (55.5 M) compared to the ester (10^{-3}–10^{-2}M), the system obeys the first-order rate law (4.2), and the reaction appears to proceed as if it were a first-order reaction. Those reactions in the cell that involve hydration, dehydration, or hydrolysis are pseudo-first order.

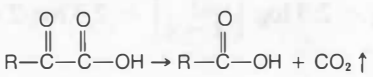
The rate expression for the *zero-order* reaction is $v = k_0$. Notice that there is no concentration term for reactants; therefore, the addi-

tion of more reactant does not augment the rate. The disappearance of reactant or the appearance of product proceeds at a constant velocity irrespective of reactant concentration. The units of the rate constant are concentration per unit time. Zero-order reaction conditions only occur in catalyzed reactions where the concentration of reactants is large enough to saturate all the catalytic sites. Under these conditions the catalyst is operating at maximum velocity, and all catalytic sites are filled; therefore, addition of more reactant cannot increase the rate.

Reversibility of Reactions

Although most chemical reactions are reversible, some directionality may be imposed on particular steps in a metabolic pathway through rapid removal of the end product by subsequent reactions in the pathway.

In the case of decarboxylation reactions



where carbon dioxide is liberated, the reaction is irreversible from a practical standpoint because CO_2 is a gas and can diffuse away from the reaction site; therefore the reaction proceeds in the forward direction by mass action. Many ligase reactions involving the nucleoside triphosphates result in release of pyrophosphate. These reactions are rendered irreversible by the hydrolysis of the pyrophosphate to 2 mol inorganic phosphate, P_i.

Schematically,

$$A + B + ATP \rightarrow A{-}B + AMP + P{-}P$$
$$P{-}P + H_2O \rightarrow 2P_i$$

The conversion of the "high-energy" pyrophosphate to inorganic phosphate imposes irreversibility on the system by virtue of the thermodynamic stability of the products.

For those reactions that are reversible, the equilibrium constant for the reaction

$$A + B \overset{K_{eq}}{\rightleftharpoons} C$$

is

$$K_{eq} = \frac{[C]}{[A][B]} \tag{4.7}$$

and can be expressed in terms of the rate constants of the forward and reverse reactions:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$$

where

$$\frac{k_1}{k_2} = K_{eq} \tag{4.8}$$

Equation (4.8) shows the relationship between thermodynamic and kinetic quantities. K_{eq} is a thermodynamic expression of the state of the system, while k_1 and k_2 are kinetic expressions that are related to the speed at which that state is reached.

Enzyme Kinetics

Terminology

There are several terms used by biochemists that are unique to enzyme kinetics. These terms all relate to defining the velocity in terms of the amount of enzyme present. The *specific activity* is defined as the micromoles of substrate converted to products per minute per milligram of enzyme protein. The specific activity indicates how fast 1 mg of an enzyme will convert 1 μmol of substrate to products. *The turnover number* is very similar to specific activity, but rather than expressing the rate of the reaction in terms of milligrams of protein, the rate is expressed in terms of moles of pure enzyme. In those cases where there is more than one catalytic center per mole of enzyme, the term *katal* is more appropriate. Katal is the number of micromoles of substrate converted to products per minute per micromole of enzyme active site.

The turnover number or katal allows a direct comparison of relative catalytic ability between enzymes. For example, the turnover numbers for catalase and α'-amylase are 5×10^6 and 1.9×10^4, respectively, indicating that catalase is about 2,500 times more active than amylase.

The *maximum velocity* V_{max} is the velocity obtained under conditions of substrate saturation of the enzyme under a given set of conditions of pH, temperature, and ionic strength.

Interaction of Enzyme and Substrate

The initial velocity of an enzyme-catalyzed reaction is dependent on the concentration of substrate as shown in Figure 4.1. As the concentration of substrate is increased (S_1–S_4), the initial velocity increases until the enzyme is completely saturated with the substrate. If one plots the initial velocities obtained at given substrate concentrations (Figure 4.2), a rectangular hyperbola is obtained. The same type of curve will be obtained for the binding of oxygen to myoglobin as a function of increasing oxygen pressure. In general, the rectangular

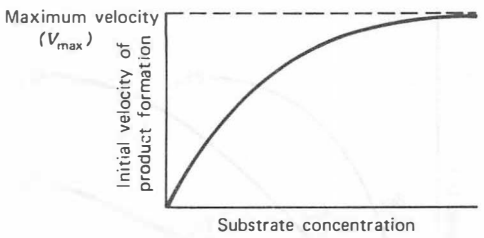


Figure 4.2
Velocity vs substrate plot for an enzyme-catalyzed reaction.
Initial velocities are plotted against the substrate concentration at which they were determined. The curve is a rectangular hyperbola, which asymptotically approaches the maximum velocity possible with a given amount of enzyme.

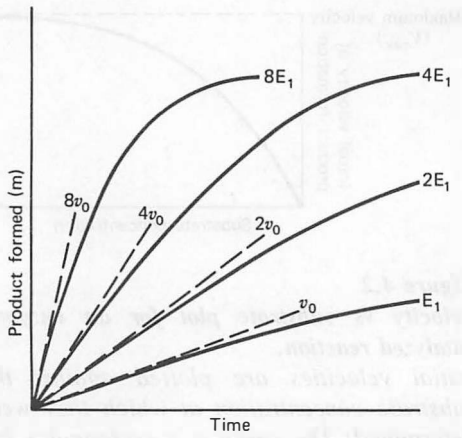


Figure 4.3
Progress curves at variable concentrations of enzyme and saturating levels of substrate.
The initial velocity doubles as the enzyme concentration doubles. Since the substrate concentrations are the same, the final equilibrium concentrations of product will be identical in each case; however, equilibrium will be reached at a slower rate in those assays containing small amounts of enzyme.

hyperbola will be obtained for any process that involves an interaction or binding of reactants or other substances at a specific but limiting number of sites. The velocity of the reaction reaches a limiting maximum at the point at which all the available sites are saturated. The curve in Figure 4.2 is referred to as the *substrate saturation curve* of an enzyme-catalyzed reaction and reflects the fact that the enzyme has a specific binding site for the substrate. Obviously the enzyme and substrate must interact in some way if the substrate is to be converted to products. Initially there must be formation of a complex between the enzyme and substrate:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \qquad (4.9)$$

The rate constant for formation of this ES complex is defined as k_1, and the rate constant for disassociation of the ES complex is defined as k_2. So far, we have described only an equilibrium binding of enzyme and substrate. The actual chemical event in which bonds are made or broken occurs in the ES complex. The conversion of substrate to products then occurs from the ES complex with a rate constant k_3. Therefore, equation (4.9) is transformed to

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + \text{products} \qquad (4.10)$$

Equation (4.10) is a general statement of the mechanism of enzyme action. The equilibrium between E and S can be expressed as an affinity constant, K_a, only if the rate of the chemical phase of the reaction, k_3, is small compared to k_2; then $K_a = k_1/k_2$. We earlier used K_{eq} to describe chemical reactions. In enzymology the association or affinity constant K_a is preferred.

The initial velocity of an enzyme-catalyzed reaction is not only dependent on the amount of substrate present, but also on the enzyme concentration. Figure 4.3 shows progress curves for increasing concentrations of enzyme, where there is enough substrate to saturate the enzyme at all levels. The initial velocity doubles as the concentration of enzyme doubles. At the lower concentrations of enzyme, equilibrium is reached more slowly than at higher concentrations, but the final equilibrium position is the same.

From our discussion thus far, we can conclude that the velocity of an enzyme reaction is dependent upon both substrate and enzyme concentration.

Formulation of the Michaelis–Menten Equation

It should be recalled that in the discussion of chemical kinetics, rate equations were developed so that the velocity of the reaction could be expressed in terms of the substrate concentration. This philoso-

phy also holds for enzyme-catalyzed reactions where the ultimate goal is to develop a relationship that will allow the velocity of a reaction to be correlated with the amount of enzyme present in a biological fluid. First, a rate equation must be developed that will relate the velocity of the reaction to the substrate concentration.

In the development of this rate equation, which is known as the *Michaelis–Menten* equation, three basic assumptions are made. The first is that the ES complex is in a steady state. That is, during the initial phases of the reaction, the concentration of the ES complex remains constant, even though many molecules of substrate are converted to products via the ES complex. The second assumption is that under saturating conditions all of the enzyme is converted to the ES complex, and none is free. The third assumption is that if all the enzyme is in the ES complex, then the rate of formation of products will be the maximum rate possible, that is,

$$V_{max} = k_3[ES] \tag{4.11}$$

If one then writes the equilibrium expression for the formation and breakdown of the ES complex as

$$K_m = \frac{k_2 + k_3}{k_1} \tag{4.12}$$

then the rate expression can be obtained after suitable algebraic manipulation as

$$\text{velocity} = v = \frac{V_{max} \cdot [S]}{K_m + [S]} \tag{4.13}$$

The complete derivation of this equation is at the end of this section. The two constants in this rate equation, V_{max} and K_m, are unique to each enzyme under specific conditions of pH and temperature. For those enzymes in which $k_3 \ll k_2$, K_m becomes the reciprocal of the enzyme–substrate binding constant, that is,

$$K_m = \frac{1}{K_a}$$

and the V_{max} reflects the catalytic phase of the enzyme mechanism as suggested by equation (4.11). In other words, in this simple model the activity of the enzyme can be separated into two phases: binding of substrate followed by chemical modification of the substrate. This biphasic nature of enzyme mechanism is reinforced in the clinical example discussed in Clin. Corr. 4.1.

CLIN. CORR. 4.1 A CASE OF GOUT DEMONSTRATES TWO PHASES TO THE MECHANISM OF ENZYME ACTION

The partitioning of the Michaelis–Menten model of enzyme action into two phases-binding, followed by chemical modification of substrate, is illustrated by studies on a family with hyperuricemia and gout. The patient excreted three times the normal amount of uric acid per day and had markedly increased levels of 5′-phosphoribosylpyrophosphate (PPRP) in his red blood cells. Assays in vivo revealed that his red cell PPRP synthase activity was increased threefold. The pH optimum and the K_m of the enzyme for ATP and ribose 5-phosphate was normal, but V_{max} was increased threefold! This increase was not because of an increase in the amount of enzyme; immunologic testing with a specific antibody to the enzyme revealed similar titers in the normal and in the patient's red cells. This finding demonstrates that the binding of substrate as reflected by K_m and the subsequent chemical event in catalysis, which is reflected in V_{max}, are separate phases of the overall catalytic process. This situation holds only for those enzyme mechanisms in which $k_2 \gg k_3$.

Notice that if one allows the initial velocity, v_0, to be equal to $\frac{1}{2}V_{max}$ in equation (4.13), K_m will become equal to [S]:

$$\tfrac{1}{2}V_{max} = \frac{V_{max} \cdot [S]}{K_m + [S]}$$

$$K_m + [S] = \frac{2V_{max} \cdot [S]}{V_{max}}$$

$$K_m = S$$

Therefore, from a substrate saturation curve the numerical value of the K_m can be derived by graphical analysis, as shown in Figure 4.4. In other words, the K_m is equal to the substrate concentration that will give half the maximum velocity.

Linear Form of the Michaelis–Menten Equation

In practice the determination of K_m from the substrate saturation curve is not very accurate, because V_{max} is approached asymptotically. If one takes the reciprocal of equation (4.13) and separates the variables into a format consistent with the equation of a straight line ($y = mx + b$), then

$$\frac{1}{v_0} = \frac{K_m}{V_{max}} \times \frac{1}{[S]} + \frac{1}{V_{max}}$$

A plot of the reciprocal of the initial velocity vs the reciprocal of the initial substrate concentration yields a line whose slope is K_m/V_{max} and whose y intercept is $1/V_{max}$. Such a plot is shown in Figure 4.5. It is often easier to obtain the K_m from the intercept on the X axis, which is $-1/K_m$.

This linear form of the Michaelis–Menten equation is often referred to as the *Lineweaver–Burk* or double reciprocal plot. Its advantage is that statistically significant values of K_m and V_{max} can be obtained directly with six to eight data points.

Derivation of the Michaelis–Menten Equation

The generalized statement of the mechanism of enzyme action is

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P \qquad (4.10)$$

If we assume that the rate of formation of the ES complex is balanced by its rate of breakdown (the steady-state assumption),

then we can write

$$v_{\text{formation}} = k_1[\text{S}]\,[\text{E}]$$

and

$$v_{\text{breakdown}} = k_2[\text{ES}] + k_3[\text{ES}] = [\text{ES}](k_2 + k_3)$$

If we set the rate of formation equal to the rate of breakdown, then

$$k_1[\text{S}]\,[\text{E}] = [\text{ES}](k_2 + k_3)$$

After dividing both sides of the equation by k_1, we have

$$[\text{S}]\,[\text{E}] = [\text{ES}]\left[\frac{k_2 + k_3}{k_1}\right] \tag{4.14}$$

If we now define the ratio of the rate constants $(k_2 + k_3)/k_1$ as K_m, the Michaelis constant, and substitute it into equation (4.14), then

$$[\text{S}]\,[\text{E}] = [\text{ES}]K_\text{m} \tag{4.15}$$

Since [E] is equal to the free enzyme, we must express its concentration in terms of the total enzyme added to the system minus any enzyme in the [ES] complex, that is,

$$[\text{E}] = [\text{E}_\text{t} - \text{ES}]$$

Upon substitution of the equivalent expression for E into equation (4.15) we have

$$[\text{S}]\,([\text{E}_\text{t}] - [\text{ES}]) = [\text{ES}]K_\text{m}$$

Dividing through by [S] yields

$$[\text{E}_\text{t}] - [\text{ES}] = \frac{[\text{ES}]K_\text{m}}{[\text{S}]}$$

and dividing through by [ES] yields

$$\frac{[\text{E}_\text{t}]}{[\text{ES}]} - 1 = \frac{K_\text{m}}{[\text{S}]} \quad \text{or} \quad \frac{[\text{E}_\text{t}]}{[\text{ES}]} = \frac{K_\text{m}}{[\text{S}]} + 1 = \frac{K_\text{m} + [\text{S}]}{[\text{S}]} \tag{4.16}$$

We now need to obtain an alternate expression for $[\text{E}_\text{t}]/[\text{ES}]$, since

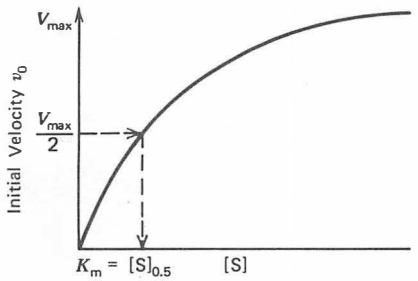


Figure 4.4
Graphic estimation of $\mathbf{K}_m$ for the v vs [S] plot. *$\mathbf{K}_m$ is the substrate concentration at which the enzyme has half-maximal activity.*

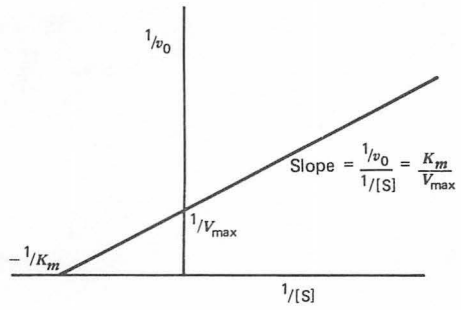


Figure 4.5
Determination of $\mathbf{K}_m$ and $\mathbf{V}_{max}$ from the Lineweaver–Burk double reciprocal plot.
Plots of the reciprocal of the initial velocity vs the reciprocal of the substrate concentration used to determine the initial velocity yield a line whose x *intercept is $-1/\mathbf{K}_m$.*

[ES] cannot be measured easily, if at all. When the enzyme is saturated with substrate all the enzyme will be in the ES complex, and none will be free ($[E_t] = [ES]$), and the velocity observed will be the maximum possible; therefore, $V_{max} = k_3[E_t]$ [See equation (4.11).] When $[E_t]$ is not equal to [ES], $v = k_3[ES]$. From these two expressions we can obtain the ratio of $[E_t]/[ES]$, that is,

$$\frac{[E_t]}{[ES]} = \frac{V_{max}/k_3}{v/k_3} = \frac{V_{max}}{v} \tag{4.17}$$

Substituting this value of $[E_t]/[ES]$ into equation (4.16) yields the Michaelis–Menten equation:

$$\frac{V_{max}}{v} = \frac{K_m + [S]}{[S]}$$

or

$$v = \frac{V_{max}[S]}{K_m + [S]}$$

Enzyme-Catalyzed Reversible Reactions

As has been indicated previously, enzymes do not alter the equilibrium constant of a reaction; consequently, in a reaction

$$S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P$$

the direction of flow of material, either in the forward direction or the reverse direction, will depend on concentration of S relative to P and the equilibrium constant of the reaction. Since enzymes catalyze the forward as well as the reverse reaction, a problem may arise if the product has an affinity for the enzyme, which is similar to that of the substrate. In this case the product can easily rebind to the active site of the enzyme and will compete with the substrate for that site. In such cases the product inhibits the reaction as the concentration of product increases. The Lineweaver–Burk plot will not be linear in those cases where the enzyme is susceptible to product inhibition. If the subsequent enzyme in the metabolic pathway has great affinity for the product and removes it, then product inhibition may not occur.

Product inhibition in a metabolic pathway provides a limited means of controlling or modulating the flux of substrates through the pathway. As the end product of the pathway increases, each intermediate will also increase via mass action. If one or more enzymes in the pathway are particularly sensitive to product inhibition, the output of the end product of the pathway will be suppressed.

Reversibility of a pathway or a particular enzyme-catalyzed reaction is dependent upon the rate of product removal and the turnover numbers of the enzymes in the pathway. If the enzymes in the pathway have high turnover numbers and the end product is a gas or is quickly removed, then the pathway may be physiological unidirectional.

Multisubstrate Reactions

Most enzymes utilize more than one substrate, or they act upon one substrate plus a coenzyme and generate one or more products. In any case, a K_m must be determined for each substrate and coenzyme involved in the reaction when establishing an enzyme assay.

Mechanistically, enzyme reactions are divided into two major categories, ping-pong and sequential. There are many variations on these major mechanisms.

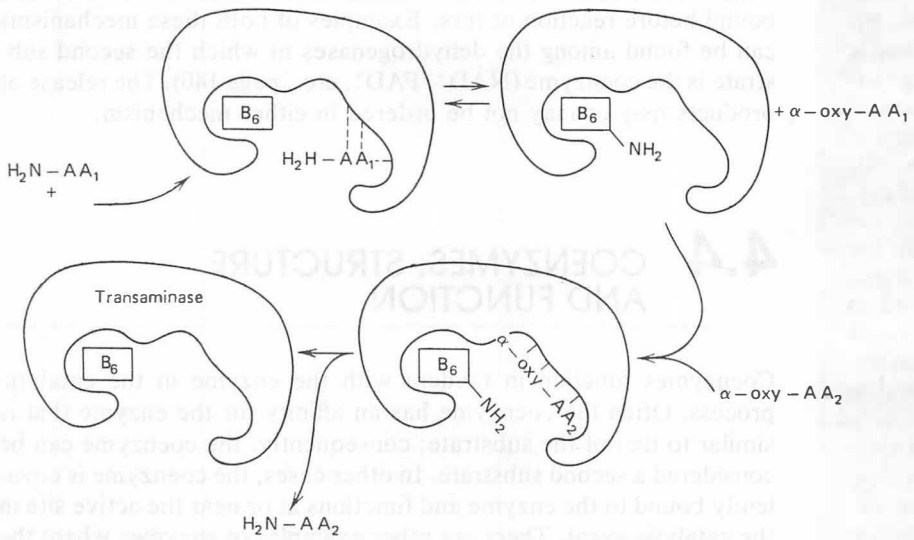


Figure 4.6
Schematic representation of the transaminase reaction mechanism—an example of a ping-pong mechanism.
Enzyme-bound vitamin B_6 coenzyme accepts the α-amino group from the first amino acid, which is then released from the enzyme as an α-keto acid. The acceptor α-keto acid is then bound to the enzyme, and the B_6-bound amino group is transferred to it, forming a new amino acid which is then released from the enzyme. The terms "oxy" and "keto" are used interchangeably.

CLIN. CORR. 4.2 MUTATION OF A COENZYME BINDING SITE RESULTS IN CLINICAL DISEASE

Cystathioninuria is a genetic disease in which the enzyme γ-cystathionase is either deficient or inactive. Cystathionase catalyzes the reaction:

Cystathionine $\longrightarrow$ cysteine + α-ketobutyrate

Deficiency of the enzyme leads to accumulation of cystathionine in the plasma and mental retardation.

Since cystathionase is a vitamin B_6-dependent enzyme, pyridoxal phosphate is the coenzyme, B_6 was administered to patients whose fibroblasts contained material that cross-reacts with antibody against cystathionase. Many respond to B_6 therapy with a fall in plasma levels of cystathionine. Such patients are making the apoenzyme. In one particular patient the activity of the enzyme was undetectable in fibroblast homogenates but increased to 31% of normal with the addition of 1 mM pyridoxal phosphate to the assay mixture. It is thought that the K_m for B_6 binding to the enzyme is increased because of a mutation of the binding site. Activity is partially restored by increasing the concentration of B_6. Apparently these patients require a higher steady-state concentration of B_6 to elicit any γ-cystathionase activity.

The *ping-pong mechanism* can be diagrammatically outlined as follows:

$$E + A \longrightarrow EA \xrightarrow{P_1\uparrow} E' \xrightarrow{B\downarrow} E'B \longrightarrow P_2 + E$$

in which substrate A is converted to product P_1, which is released before the second substrate B will bind to the modified enzyme E′. B is then converted to product P_2 and the enzyme regenerated. A good example of this mechanism is the transaminase catalyzed reaction (page 546) in which the α-amino group of amino $acid_1$ (A) is transferred to the enzyme and the newly formed keto $acid_1$ is released (P_1) followed by the binding of the acceptor keto $acid_2$ (B) and release of amino $acid_2$ (P_2). This reaction is schematically outlined in Figure 4.6.

In the sequential mechanism, if the two substrates A and B can bind in any order, it is a *random mechanism;* if the binding of A is required before B can be bound, then it is an *ordered mechanism.* In either case the reaction is bimolecular, that is, both A and B must be bound before reaction occurs. Examples of both these mechanisms can be found among the dehydrogenases in which the second substrate is the coenzyme (NAD^+, FAD^+, etc., page 140). The release of products may or may not be ordered in either mechanism.

4.4 COENZYMES: STRUCTURE AND FUNCTION

Coenzymes function in tandem with the enzyme in the catalytic process. Often the coenzyme has an affinity for the enzyme that is similar to that of the substrate; consequently, the coenzyme can be considered a second substrate. In other cases, the coenzyme is covalently bound to the enzyme and functions at or near the active site in the catalytic event. There are other examples of enzymes where the role of the coenzyme falls between these two extremes.

Several, but not all, of the coenzymes are synthesized from the B vitamins. Vitamin B_6, pyridoxine, requires little modification to be transformed to the active coenzyme, pyridoxal phosphate (page 545). Clin. Corr. 4.2 points out the importance of the coenzyme binding site and how alterations in this site cause metabolic dysfunction.

In contrast to B_6, niacin requires major alteration by the mamma-

lian cell before it is capable of acting as a coenzyme. This metabolic interconversion is outlined on page 665.

The structure and function of only two vitamin B coenzymes, niacin and riboflavin, and of ATP will be discussed in this chapter. The structure and function of coenzyme A (page 273), thiamine (page 273), biotin, and B_{12} are included in those chapters dealing with enzymes dependent upon the given coenzyme for activity.

Adenosine Triphosphate

Adenosine triphosphate (ATP) often functions as a second substrate but can also serve as a cofactor in modulating the activity of specific enzymes. This compound is so pivotal that its structure and function will be introduced here. ATP (Figure 4.7) can be synthesized de novo in all mammalian cells.

The nitrogenous heterocyclic ring is adenine. To the adenine is affixed a ribosyl 5′-triphosphate. The functional end of the molecule is the reactive triphosphate which is shown in the ionization state found in the cell. As a cosubstrate ATP is utilized by the kinases for the transfer of the terminal phosphate to various acceptors. A typical example is the glucokinase-catalyzed reaction:

$$\text{Glucose} + \text{ATP} \longrightarrow \text{glucose 6-phosphate} + \text{ADP}$$

ADP is adenosine diphosphate. The combination of adenine plus ribose is adenosine (page 632).

ATP has an additional role, other than cosubstrate: in a number of specific enzyme reactions it serves as a modulator of the activity of the enzyme. These particular enzymes have binding sites for ATP, occupancy of which changes the affinity or reactivity of the enzyme toward its substrates. Mechanistically, ATP is acting as an allosteric effector in these cases (page 171).

Coenzymes of Niacin

Niacin is pyridine-3-carboxylic acid. It is converted to two major coenzymes which are involved in the oxidoreductase class of enzymes. These coenzymes are NAD^+, nicotinamide adenine dinucleotide, and $NADP^+$, nicotinamide adenine dinucleotide phosphate. There are dehydrogenases that function with $NADP^+$ as coenzyme but not with NAD^+. The reverse is also true. Such an arrangement allows for specificity and control over dehydrogenases that reside in the same subcellular compartment.

Structurally, NAD^+ is composed of adenosine and *N*-ribosyl-

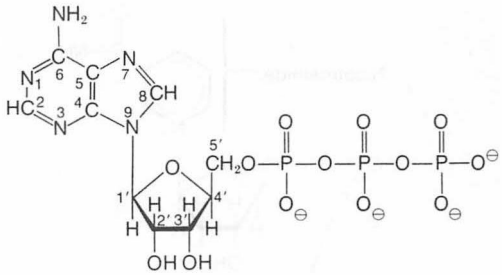


Figure 4.7
Adenosine triphosphate (ATP).

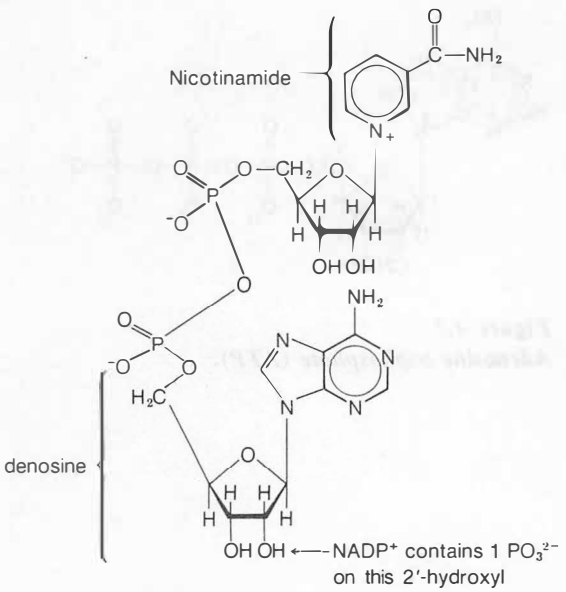


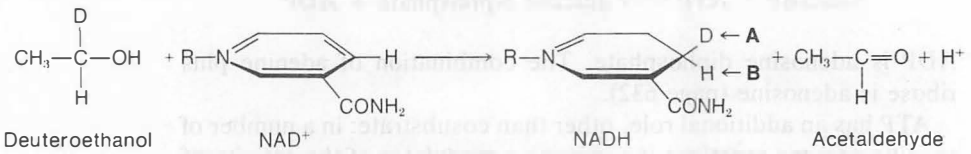
Figure 4.8
Nicotinamide adenine dinucleotide (NAD).

nicotinamide linked through a pyrophosphate linkage between the 5′-hydroxyls of the two ribosyl moieties (Figure 4.8). $NADP^+$ differs structurally from NAD in having an additional phosphate esterified to the 2′-hydroxyl of the adenosine moiety.

Both coenzymes function as intermediates in the transfer of two electrons between an electron donor and acceptor. The donor and acceptor need not be involved in the same metabolic pathway. In other words, the reduced form of these nucleotides acts as a common "pool" of electrons that arise from many oxidative reactions and can be used for various reductive reactions.

The adenine, ribose, and pyrophosphate components of NAD^+ are involved in the binding of NAD^+ to the enzyme. Enzymes requiring $NADP^+$ have a specific cationic region in their $NADP^+$ binding site that is positioned so as to form an ionic bond with the 2′-phosphate of $NADP^+$. This enhances binding of $NADP^+$ in preference to NAD^+ in these particular enzymes.

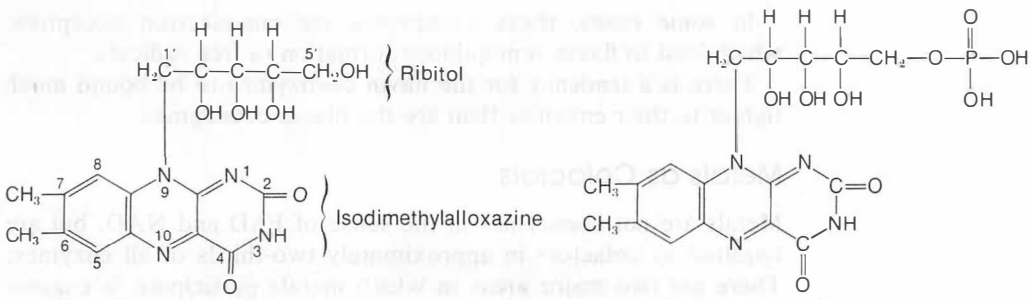
The nicotinamide portion of the molecule is involved in reversibly accepting and donating two electrons at a time. It is the active center of the coenzyme. In the oxidation of deuterated ethanol by alcohol dehydrogenase, NAD^+ accepts two electrons and one deuterium from the ethanol, and the other hydrogen is released as a proton.



The specific binding of NAD^+ to the enzyme surface confers a chemically recognizable "topside" and "bottom side" to the planar nicotinamide. The former is known as the **A** face and the latter the **B** face. In the case of alcohol dehydrogenase, the proton or the deuterium which serves as a tracer and two electrons are added to the **A** face. Other dehydrogenases utilize the **B** face. The particular effect just described demonstrates how enzymes are able to induce specificity into chemical reactions by virtue of the asymmetric binding of coenzymes and substrates.

Coenzymes of Riboflavin

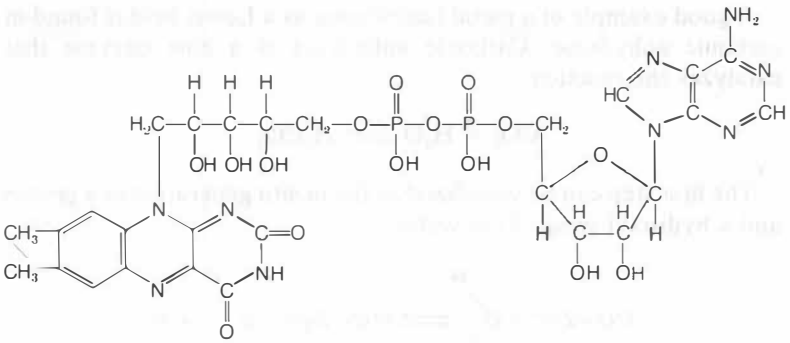
The two coenzyme forms of riboflavin are FMN, riboflavin 5′-phosphate, and FAD (flavin adenine dinucleotide). The vitamin riboflavin consists of the heterocyclic ring, isoalloxazine (flavin) connected through N (9) to the alcohol ribitol as shown.



Riboflavin

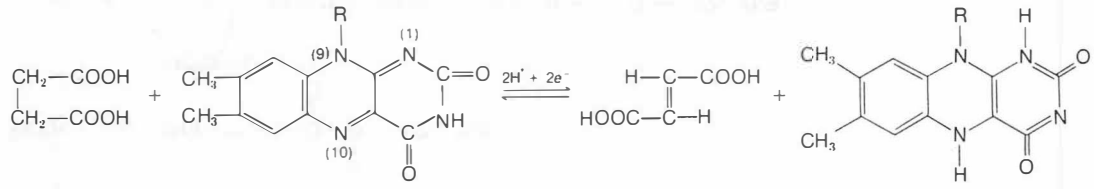
Riboflavin monophosphate (FMN)

FMN has a phosphate esterified to the 5′-hydroxyl of riboflavin. FAD is structurally analogous to NAD in having adenosine linked through a pyrophosphate linkage to a heterocyclic ring, in this case riboflavin.



Flavin adenine dinucleotide (FAD)

FAD and FMN function in oxidation–reduction reactions by accepting and donating two electrons through the isoalloxazine ring. A typical example of FAD participation in an enzyme reaction is the oxidation of succinate to fumarate by succinic dehydrogenase (page 285).



Succinic Acid FAD Fumaric acid $FADH_2$

In some cases, these coenzymes are one-electron acceptors, which lead to flavin semiquinone formation (a free radical).

There is a tendency for the flavin coenzymes to be bound much tighter to their enzymes than are the niacin coenzymes.

Metals as Cofactors

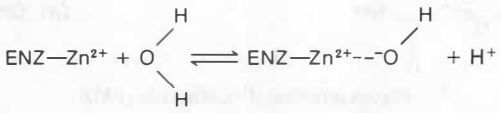
Metals are not coenzymes in the sense of FAD and NAD, but are required as cofactors in approximately two-thirds of all enzymes. There are two major areas in which metals participate in enzyme reactions—through their ability to act as *Lewis acids* and through various modes of *chelate* formation. Chelates are organometallic coordination complexes. A good example of a chelate is the complex between iron and porphyrin to form a heme (page 309).

Those metals that act as Lewis acid catalysts are found among the transition metals like Zn, Fe, Mn, and Cu, which have empty *d* electron orbitals that can act as electron sinks. The alkaline earth metals such as K and Na do not possess this ability.

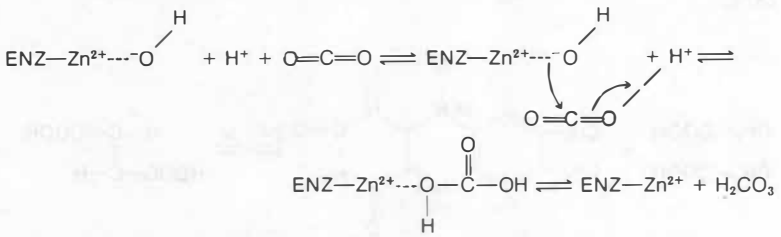
A good example of a metal functioning as a Lewis acid is found in carbonic anhydrase. Carbonic anhydrase is a zinc enzyme that catalyzes the reaction

$$CO_2 + H_2O \rightleftharpoons H_2CO_3$$

The first step can be visualized as the in situ generation of a proton and a hydroxyl group from water:



The proton and hydroxyl are subsequently added to the carbon dioxide and carbonic acid is released. The reactions are presented in a stepwise fashion for clarity. Actually, the reactions may occur in a concerted fashion, that is, all at one time.



The metal can also promote catalysis by binding substrate at the site of the bond cleavage. In carboxypeptidase, the carbonyl oxygen

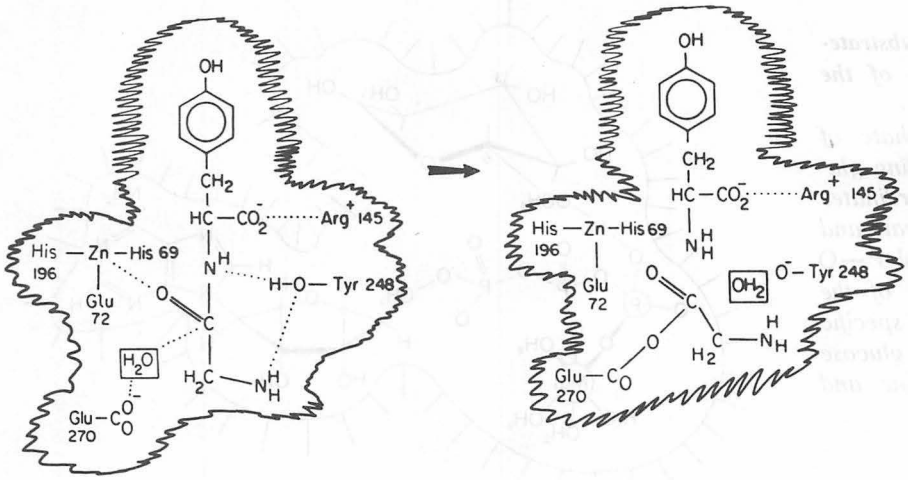


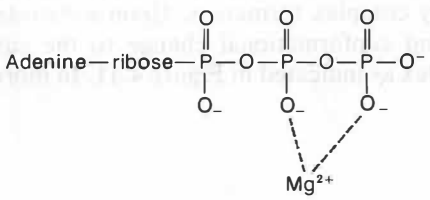
Figure 4.9
The role of zinc in carboxypeptidase A.
Enzyme-bound zinc polarizes the peptide carbonyl making the carbonyl carbon more positive and susceptible to nucleophilic attack by glutamic acid residue 270 in the active site. The end result is cleavage of the peptide bond, release of a new amino-terminal peptide and covalent addition of the remainder of the substrate to the enzyme through anhydride linkage. The latter is then released by water hydrolysis.
From W. Lipscomb, J. Hartsuck, F. Qurocho and G. Reeke, Jr., *Proc. Natl. Acad. Sci. USA*, 64:39, 1969. Reprinted with permission.

atom is chelated to the zinc. The resulting flow of electrons from the carbonyl carbon to the electropositive metal increases the susceptibility of the peptide bond to cleavage by nucleophiles such as water or carboxylate. This is schematically shown in Figure 4.9.

Role of the Metal as a Structural Element

The functioning of a metal as a Lewis acid requires chelate formation. In addition, various modes of chelation occur between metal, enzyme, and substrate that are structural in nature, but in which no acid catalysis occurs.

In several of the kinases, creatine kinase being the best example, the true substrate is not ATP but Mg^{2+}-ATP.



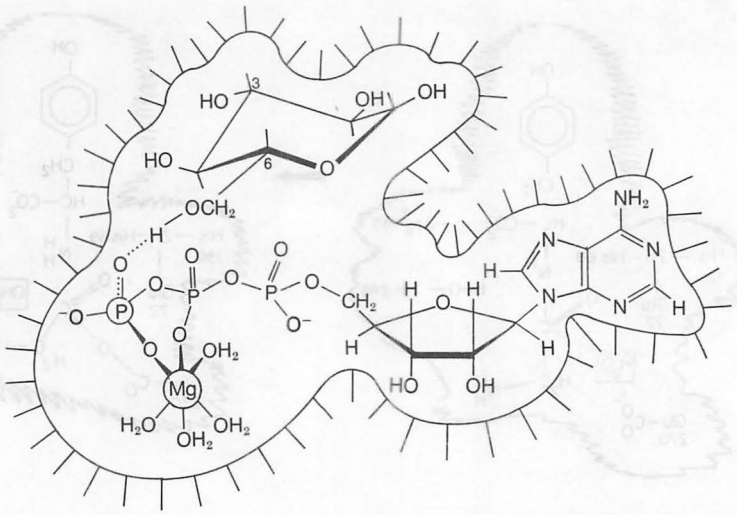


Figure 4.10
Model of the role of magnesium as a substrate-bridged complex in the active site of the kinases.
In hexokinase the terminal phosphate of ATP is transferred to glucose, yielding glucose 6-phosphate. Magnesium coordinates with the ATP to form the true substrate and in addition may labilize the terminal P—O bond of ATP to facilitate transfer of the phosphate to glucose. There are specific binding sites in the active site for glucose (upper left) as well as the adenine and ribose moieties of ATP.

In this case, the magnesium does not interact directly with the enzyme. It may serve to neutralize the negative charge density on ATP and facilitate binding to the enzyme. Ternary complexes of this configuration are known as "substrate-bridged" complexes and can be schematically represented as Enz—S—M. A hypothetical scheme for the binding of Mg–ATP and glucose in the active site of hexokinase is shown in Figure 4.10. All the kinases except muscle pyruvate kinase and phospho*enol*pyruvate carboxykinase are substrate-bridged complexes.

In pyruvate kinase the metal Mg^{2+} serves to chelate the ATP to the enzyme as shown in Figure 4.11. The absence of the metal cofactor results in failure of the ATP to bind to the enzyme. Enzymes of this class are "metal-bridged" ternary complexes, Enz—M—S. All metalloenzymes are of this type. *Metalloenzymes* are enzymes containing a tightly bound transition metal such as Zn^{2+} and Fe^{2+}. Several enzymes catalyzing enolization and elimination reactions are metal-bridged complexes.

In addition to the role of binding enzyme and substrate, metals may also bind directly to the enzyme to stabilize it in the active conformation or perhaps to induce the formation of a binding site or active site. Not only do the strongly chelated metals like Mn^{2+} play a role in this regard, but the weakly bound alkali metals (Na^+, K^+) are also important. In pyruvate kinase, K^+ has been found to induce an initial conformation change, which is necessary, but not sufficient, for the ternary complex formation. Upon substrate binding, K^+ induces a second conformational change to the catalytically active ternary complex as indicated in Figure 4.11. In more general terms it

is thought that Na^+ and K^+ stabilize the active conformation of the enzyme, but are passive from the catalytic standpoint.

Role of Metals in Oxidation and Reduction

The iron-sulfur enzymes, often referred to as *nonheme iron* proteins, are a unique class of metalloenzymes in which the active center consists of one or more clusters of sulfur-bridged iron chelates. These are of greater prominence in bacterial and plant systems than in mammalian cells. In mammalian systems succinic dehydrogenase (page 285), NADH dehydrogenase, and adrenodoxin are good representatives of this group of proteins. The structure of the iron chelate in these nonheme iron proteins is represented in Figure 4.12.

In these proteins the bridging sulfide is released as H_2S on acidifying the enzyme. Cysteine thiol groups from the enzyme hold the bridged binuclear iron complex in the enzyme. These particular enzymes have reasonably low reducing potentials (E_0') and function in electron transfer reactions. Adrenodoxin functions in the activation of oxygen in the steroid 11β- and 18-hydroxylases as a cosubstrate. It is not an enzyme.

The cytochromes, which are heme iron proteins, also function as cosubstrates for their respective reductases (page 308). The iron in the hemes of the cytochromes undergoes reversible one-electron

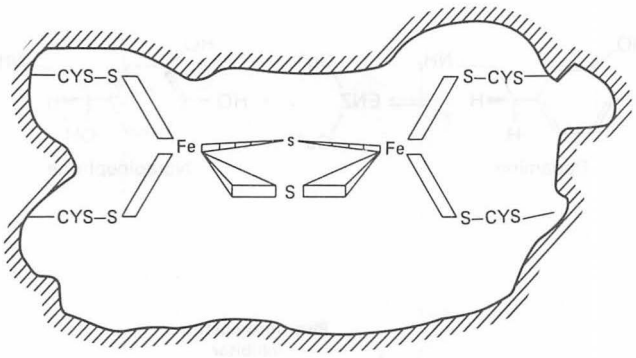


Figure 4.12
The iron binding site of adrenodoxin.
Two iron atoms are chelated to the protein via cysteine sulfhydryl groups. The two iron atoms are bridged by sulfides, which are released as hydrogen sulfide upon acidification of the protein. A formal valence state cannot be assigned to the iron because they are magnetically coupled.

Reprinted, with permission, from H. Sigel, ed., *Metal Ions in Biological Systems*, vol. 7, 1979 Marcel Dekker, Inc.

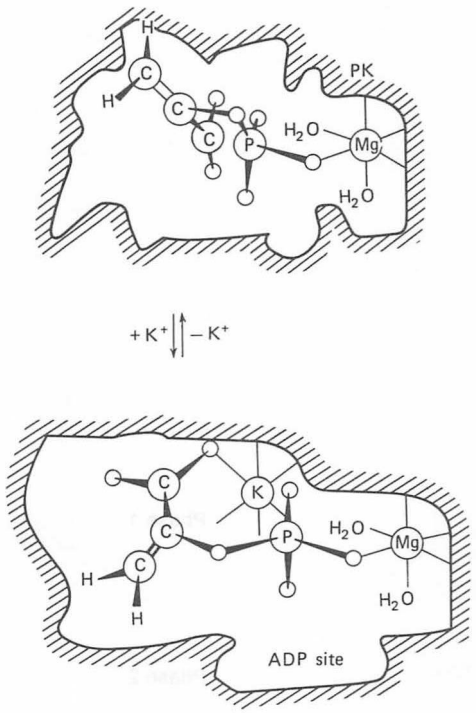


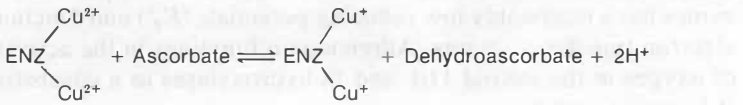
Figure 4.11
Model of the role of potassium ion in the active site of pyruvate kinase.
*Pyruvate kinase catalyzes the reaction: phospho*enol*pyruvate + ADP → ATP + pyruvate. Initial binding of K^+ induces conformational changes in the kinase, which result in increased affinity for phospho*enol*pyruvate. In addition, K^+ orients the phospho*enol*pyruvate in the correct position for transfer of its phosphate to the second substrate, which is ADP. Magnesium coordinates the substrate to the enzyme active site.*

Modified, with permission, from A. S. Mildvan, *Ann. Rev. Biochem.*, 43:365, 1974. © Annual Reviews, Inc.

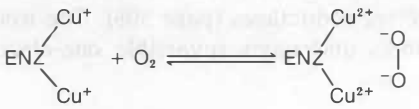
transfers. In addition, the heme is bound to the enzyme through coordination of an enzyme amino acid side chain to the iron of the heme. Thus, in the cytochromes the metal serves not only a structural role, but also participates in the chemical event.

The last role that we will discuss here is the role of metals, specifically copper and iron, in activation of molecular oxygen. Copper is an active participant in several oxidase and hydroxylase enzymes. For example, dopamine β-hydroxylase catalyzes the introduction of one oxygen atom from O_2 into dopamine to form norepinephrine. It is thought that the active form of the enzyme contains two atoms of cuprous ion, which react with O_2 to form the reactive species O_2^{2-}, which then attacks the dopamine. In other metalloenzymes other species of "active" oxygen are generated.

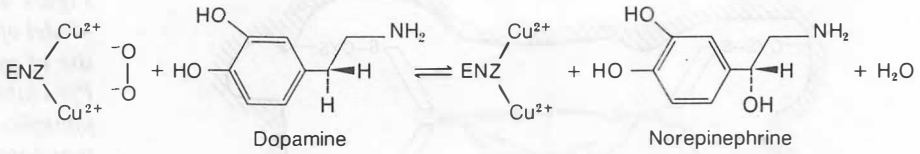
Phase 1



Phase 2



Phase 3



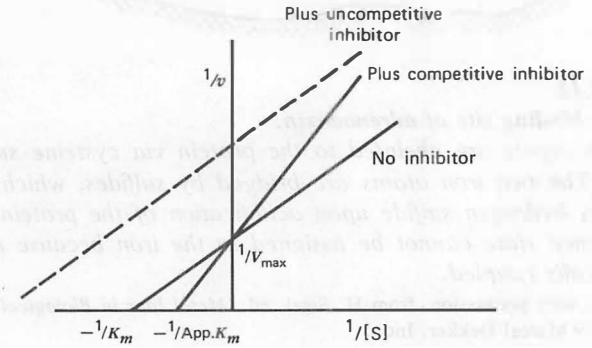


Figure 4.13
Double reciprocal plot for an enzyme subject to competitive inhibition.
A competitive inhibitor binds at the substrate binding site and effectively increases the K_m *for the substrate.*

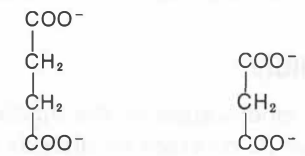
4.5 INHIBITION OF ENZYMES

Mention was made earlier of product inhibition of enzyme activity and how an entire pathway could be controlled or modulated by this mechanism (page 154). In addition to inhibition by the immediate product, products of other enzymes can also inhibit or even activate a particular enzyme. Much of current drug therapy is based on inhibition of specific enzymes with a substrate analog. Therefore, it is important to discuss inhibition in more detail. Basically, there are three major classes of inhibitors: competitive, noncompetitive, and uncompetitive.

Competitive Inhibitors

Competitive inhibitors are defined as inhibitors whose action can be reversed by increasing amounts of substrate. Competitive inhibitors are usually structurally enough like the substrate that they bind at the substrate binding site and compete with the substrate for the enzyme. Once bound, the enzyme cannot convert the inhibitor to products. Increasing substrate concentrations will displace the reversibly bound inhibitor by the law of mass action. A competitive inhibitor need not be structurally related to the substrate.

In the succinate dehydrogenase reaction, malonate is structurally similar to succinate and is a competitive inhibitor.



Succinate Malonate

Since the substrate and inhibitor are competing for the same site on the enzyme, the K_m for the substrate shows an apparent increase in the presence of inhibitor. This can be seen in a double-reciprocal plot as a shift in the X intercept ($-1/K_m$) and in the slope of the line (K_m/V_{max}). If we first establish the velocity at several levels of substrate and then repeat the experiment with a given but constant amount of inhibitor at various substrate levels, two different straight lines will be obtained as shown in Figure 4.13. As can be seen, the V_{max} does not change; hence the intercept on the Y axis remains the same. In the presence of inhibitor, the X intercept is no longer the negative reciprocal of the true K_m, but is $-1/K_{m_{app}}$ where

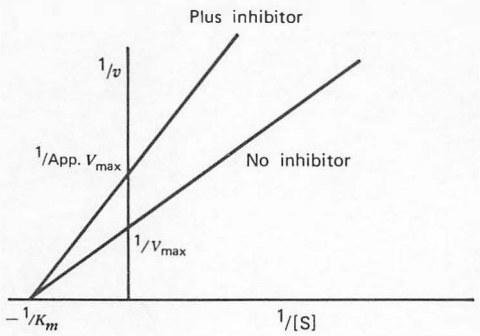


Figure 4.14
Double reciprocal plot for an enzyme subject to reversible noncompetitive inhibition.
A noncompetitive inhibitor binds at a site other than the substrate binding site; therefore, the effective K_m *does not change, but the apparent* V_{max} *decreases.*

$$K_{m_{app}} = K_m \cdot \left[1 + \frac{[I]}{K_I}\right]$$

Thus the inhibitor constant, K_I, can be determined from the concentration of inhibitor [I] used and the K_m, which was obtained from the X intercept of the line obtained in the absence of inhibitor.

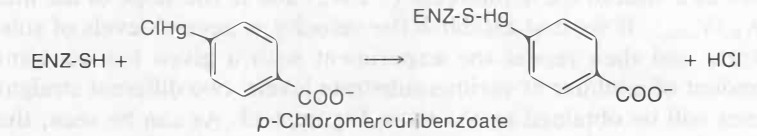
Noncompetitive Inhibitors

A noncompetitive inhibitor binds at a site other than the substrate binding site. Its inhibition is not reversed by increasing concentrations of substrate. Both binary, EI, and ternary, EIS, complexes form, both of which are catalytically inactive and are therefore, dead-end complexes. The noncompetitive inhibitor behaves as though it were removing active enzyme from the solution, resulting in a decrease in V_{max}. This effect is seen graphically in the double-reciprocal plot (Figure 4.14), where K_m does not change but V_{max} does change. Inhibition can often be reversed by exhaustive dialysis of the inhibited enzyme provided that the inhibitor has not reacted covalently with the enzyme. This case is considered under the irreversible inhibitors and is discussed below.

The uncompetitive inhibitor binds only with the ES form of the enzyme in the case of a one-substrate enzyme. The result is an apparent equivalent change in K_m and V_{max}, which is reflected in the double reciprocal plot as a line parallel to that of the uninhibited enzyme (Figure 4.13). In the case of multisubstrate enzymes the interpretation is complex and will not be considered further.

Irreversible Inhibitors

In cases of covalent modification of the binding site or the active site, inhibition will not be reversed by dialysis unless the linkage is chemically labile like an ester or thioester. The active site thiol in glyceraldehyde 3-phosphate dehydrogenase reacts with *p*-chloromercuribenzoate to form a covalent mercuribenzoate adduct of the enzyme.



Such adducts are not reversed by dialysis or substrate. Double re-

ciprocal plots show the pattern for noncompetitive inhibition (Figure 4.14).

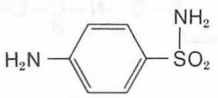
Drugs and Enzyme Inhibition

Most if not all of modern drug therapy is based on the concepts of enzyme inhibition that were covered in the previous section.

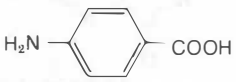
Drugs are designed with a view toward inhibiting a specific enzyme in a specific metabolic pathway. This application is most easily appreciated with the antiviral, antibacterial, and antitumor drugs, which are administered to the patient under conditions of limited toxicity. Such toxicity to the patient is unavoidable because, with the exception of cell wall biosynthesis in bacteria, there are no critical metabolic pathways that are unique to tumors, viruses, or bacteria. Hence, drugs that will kill these organisms will also kill the host cell. The one characteristic that can be taken advantage of is the comparatively short generation time of the undesirable organisms. They are much more sensitive to antimetabolites and in particular those that inhibit enzymes involved in replication. *Antimetabolites* are compounds with slight structural deviation from the natural substrate. In the chapters on metabolism, numerous examples of antimetabolites will be brought to your attention. Here we will present only a few examples which illustrate the concept.

Sulfa Drugs

Modern chemotherapy had its beginning in these compounds whose general formula is R—SO_2—NHR′. Sulfanilamide is the simplest member of the class and is an antibacterial agent because of its competition with *p*-aminobenzoic acid, which is required for bacterial growth.



Sulfanilamide



p-Aminobenzoic Acid

It is now known that bacteria cannot absorb folic acid, a required vitamin, from the host, but must synthesize it. The synthesis of folate involves the series of reactions shown in Figure 4.15.

Since sulfanilamide is a structural analog of *p*-aminobenzoate, the enzyme dihydropteroate synthetase is tricked into making a dihydropteroate containing sulfanilamide that cannot be converted to folate. Thus the bacterium is starved of the required folate and cannot grow or divide. Since man requires folate from external sources, the sulfanilamide is not harmful at the doses that will kill bacteria.

Methotrexate

The biosynthesis of purines and pyrimidines, heterocyclic bases employed in the synthesis of RNA and DNA, requires folic acid, which serves as a coenzyme in the transfer of one-carbon units from various amino acid donors (page 570). For example, in the conversion of deoxyuridine phosphate to thymidine phosphate, the methyl

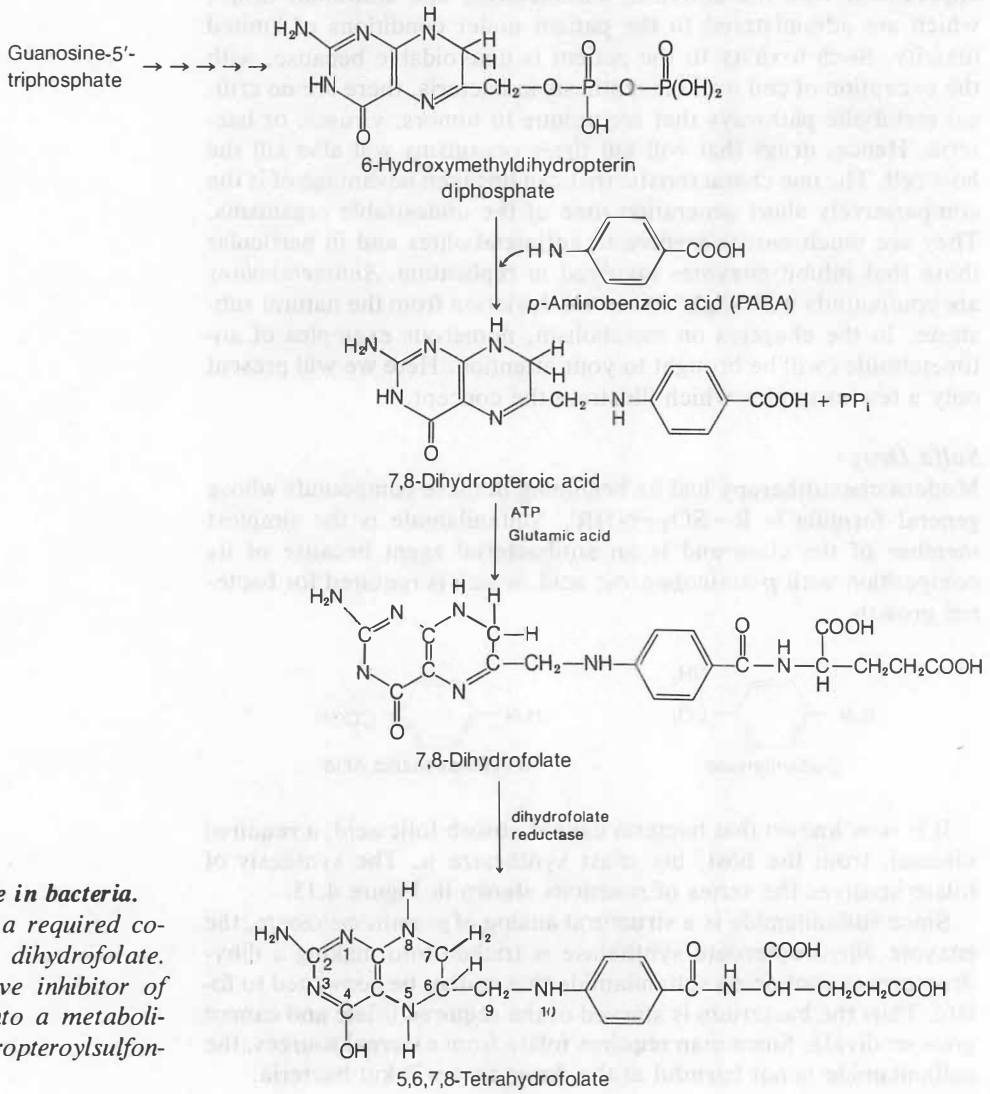
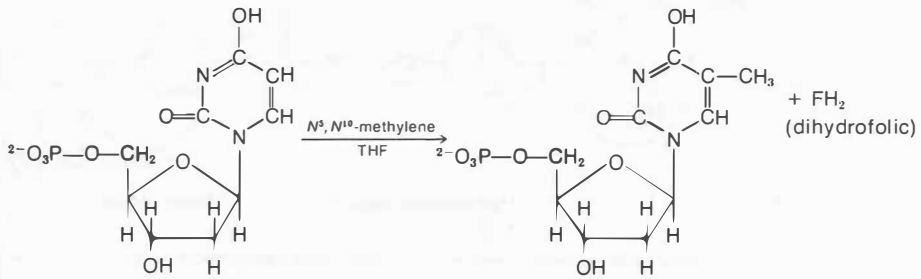


Figure 4.15
The route of synthesis of folate in bacteria. *p-Aminobenzoate (PABA) is a required co-factor for the synthesis of dihydrofolate. Sulfanilamide is a competitive inhibitor of PABA and is incorporated into a metabolically inactive 7,8-dihydropteroylsulfonamide.*

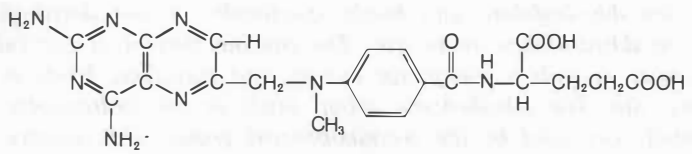
group in thymidine phosphate arises from N^5,N^{10}-methylene tetrahydrofolate.



Deoxyuridine-5'-phosphate
(Deoxyuridylic acid)

Thymidine-5'-phosphate
(thymidylic acid)

The formation of the methylene tetrahydrofolate first requires the conversion of the 7,8-dihydrofolate (Figure 4.15) to the tetrahydrofolate. This reaction is catalyzed by the dihydrofolate reductase.



Methotrexate
(4-amino-N^{10}-methyl folic acid)

Methotrexate is a structural analog of folate. It has been used with great success in childhood leukemia. Its mechanism of action is based on competition with dihydrofolate for the dihydrofolate reductase. It binds 1,000-fold better than the natural substrate and is a powerful competitive inhibitor of the enzyme. This being the case, the synthesis of thymidine nucleotide stops in the presence of methotrexate because of failure of the one-carbon metabolic system. Since cell division is dependent on thymidine as well as the other nucleotides, the leukemia cell cannot multiply. One problem is that rapidly dividing human cells such as those in bone marrow are sensitive to the drug for the same reasons. Also, prolonged usage stimulates the tumor cells to produce larger amounts of the reductase enzyme, thus becoming resistant to the drug.

Nonclassical Antimetabolites

These compounds are also known as suicide substrates or active site directed inhibitors. The compounds are constructed so that they have an affinity for the active site, but in addition have a chemically reactive group that will form a covalent adduct with a reactive amino

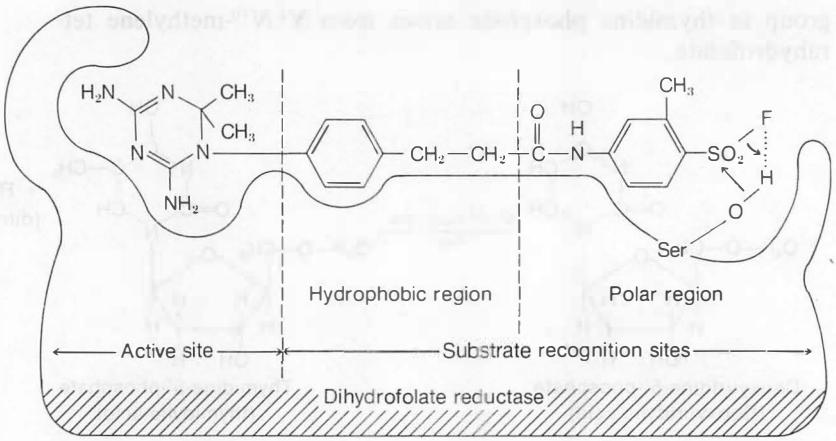


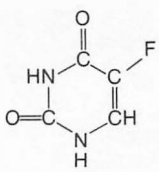
Figure 4.16
Suicide substrate inactivation of tetrahydrofolate reductase.
The suicide substrate, a substituted dihydrotriazine, structurally resembles dihydrofolate and binds specifically to the dihydrofolate site on dihydrofolate reductase. The triazine portion of the suicide substrate resembles the pterin moiety and therefore, binds in the active site. The ethylbenzene group binds in the hydrophobic site normally occupied by the p-aminobenzoyl group. The reactive end of the suicide substrate contains a reactive sulfonyl fluoride that forms a covalent linkage with a serine hydroxyl on the enzyme surface. Thus the suicide substrate irreversibly inhibits the enzyme by blocking access of dihydrofolate to the active site.

acid in the active site of the enzyme. Thus the compounds are specific for a particular enzyme and the inhibition is irreversible. For example, the compound shown in Figure 4.16 is a suicide substrate for the dihydrofolate reductase because the compound structurally resembles dihydrofolate and is specifically bound at the active site where the reactive benzylsulfonyl fluoride is positioned so as to react with a serine hydroxyl in the substrate binding site. Covalent binding of this suicide substrate to the enzyme prevents binding of the normal substrate and leads to inhibition of the enzyme.

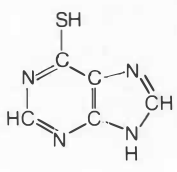
Other Antimetabolites

Two other analogs of the purines and pyrimidines will be mentioned in order to emphasize the structural similarity of chemotherapeutic agents to normal substrates.

Fluorouracil is an analog of thymine in which the ring bound methyl is substituted by fluorine. The deoxynucleotide of this compound is an irreversible inhibitor of the enzyme thymidylate synthetase.



5-Fluorouracil



6-Mercaptopurine

6-Mercaptopurine is an analog of hypoxanthine and therefore of adenine and guanine. 6-Mercaptopurine is a broad-spectrum antimetabolite because of its competition in most reactions involving adenine and guanine or their derivatives.

The antimetabolites discussed have been related to purine and pyrimidine metabolism. However, the general concepts developed here can be applied to any enzyme or metabolic pathway.

4.6 ALLOSTERIC CONTROL OF ENZYME ACTIVITY

Allostery and Cooperativity

Although the substrate site and the active site of an enzyme are well-defined structures, the activity of many enzymes can be modulated by ligands acting in ways other than competitive or noncompetitive inhibitors. A *ligand* is any molecule that is bound to a macromolecule; the term is not limited to small organic molecules such as ATP, but is extended to low molecular weight proteins. Ligands can be activators, inhibitors, or even the substrates of enzymes. Those ligands that cause a change in enzymatic activity, but are unchanged as a result of enzyme action, are referred to as *effectors, modifiers,* or *modulators.* Most of the enzymes subject to modulation by ligands are rate-determining enzymes in metabolic pathways. In order to appreciate the mechanisms of control of metabolic pathways, the principles governing the allosteric and cooperative behavior of individual enzymes must be understood.

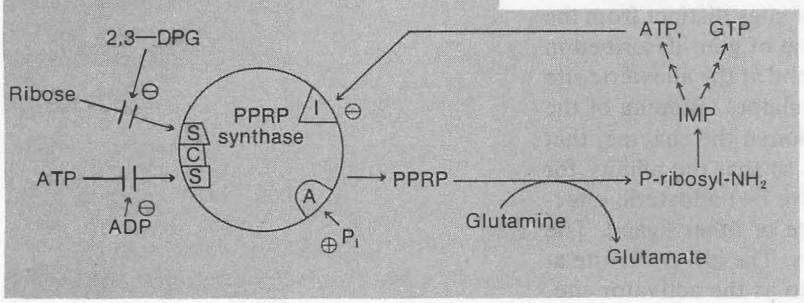
In addition to the substrate binding site and the active site, which we have previously discussed, those enzymes that respond to modulators have additional site(s) known as allosteric site(s). *Allosteric* is derived from the Greek root *allo,* meaning "the other"; hence the allosteric site is a unique region of the enzyme that is different from the substrate binding site (which is often the catalytic site). The existence of allosteric inhibitory and activator sites distinct from the substrate binding site is illustrated by the case of gout described in Clin. Corr. 4.3. The modulating ligands that bind at the allosteric site are known as the allosteric effectors or modulators. Binding of the allosteric effector causes an allosteric transition in the enzyme, that is, the conformation of the enzyme changes, so that the affinity for the substrate or other ligand changes. Positive (+) allosteric effectors increase the enzyme affinity for substrate or other ligand. The reverse is true for negative allosteric effectors. The allosteric site at which the positive effector binds is referred to as the activator site. The negative effector binds at an inhibitory site.

Allosteric enzymes are divided into two classes based on the effect of the allosteric effector on the K_m and V_{max}. In the K class the effector alters the K_m but not the V_{max}, whereas in the V class the effector alters the V_{max} but not the K_m. K class enzymes give double-reciprocal plots like those given by competitive inhibitors (Figure 4.13) and V class enzymes give double-reciprocal plots like those of noncompetitive inhibitors (Figure 4.14). However, it is inappropriate to use the terms competitive and noncompetitive with allosteric enzyme systems because the mechanism of the effect of an allosteric inhibitor on a V or K enzyme is quite different from the mechanism of a simple competitive or noncompetitive inhibitor. For example, in the K class the inhibitor binds at an allosteric site, which then affects the affinity of the substrate binding site for the substrate, whereas in simple competitive inhibition the inhibitor competes with substrate for the substrate binding site. In the V class enzymes, positive and negative allosteric modifiers increase or decrease the rate of breakdown of the ES complex to products, that is, the catalytic rate constant, k_3, is affected and not the substrate binding constant. There are a few enzymes in which both K_m and V_{max} are affected.

In theory a monomeric enzyme can undergo an allosteric transition in response to a modulating ligand. In practice only two monomeric allosteric enzymes have been found, ribonucleoside diphosphate reductase and pyruvate-UDP-*N*-acetylglucosamine transferase. Most allosteric enzymes are oligomeric, that is, they consist of several subunits. The identical subunits are designated as *protomers*. Each protomer may consist of one or more polypeptide chains. As a consequence of the oligomeric nature of allosteric enzymes, binding of ligand to one protomer can affect the binding of ligands on other protomers in the oligomer. Such ligand effects are referred to as homotropic interactions. The transmission of the homotropic effects between protomers is one aspect of cooperativity, considered

CLIN. CORR. 4.3 A CASE OF GOUT DEMONSTRATES THE DIFFERENCE BETWEEN AN ALLOSTERIC SITE AND THE SUBSTRATE BINDING SITE

The realization that allosteric inhibitory sites are separate from activator sites as well as from the substrate binding and the catalytic site is illustrated by a study of a gouty patient whose red blood cell PPRP levels are increased. It was found that the patient's PPRP synthase had a normal K_m, V_{max} and sensitivity to activation by inorganic phosphate. However, it displayed reduced sensitivity to the normal allosteric inhibitors AMP, ADP GDP, and 2,3-diphosphoglycerate (2,3-DPG). The increased PPRP levels and hyperuricemia arose because endogenous end products of the pathway (ATP, GTP) were not able to control the activity of the synthase as is normally the case. Apparently mutation has resulted in a change in the inhibitory site (I) or in the coupling mechanism between the I and the catalytic site (C).



in detail later in this chapter. Substrate influencing substrate, activator influencing activator, or inhibitor influencing inhibitor binding are homotropic interactions. Homotropic interactions are almost always positive.

A *heterotropic interaction* is defined as the effect of one ligand on the binding of a different ligand. For example, the effect of an allosteric inhibitor on the binding of substrate or the effect of an allosteric inhibitor on the binding of an allosteric activator are heterotropic interactions. Heterotropic interactions can be either positive or negative and can occur in monomeric allosteric enzymes. Both heterotropic and homotropic effects, in an oligomeric enzyme, are mediated by cooperativity between subunits.

Based on the foregoing descriptions of allosteric enzymes, two models are pictured in Figure 4.17. In panel *a* a model for a monomeric enzyme is shown, and in panel *b* a model for an

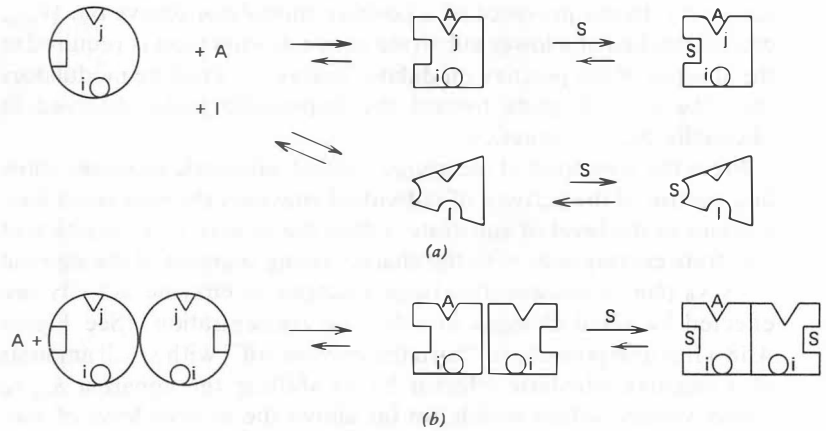


Figure 4.17
Models of allosteric enzymes systems.
(a) A model of a monomeric enzyme. Binding of a positive allosteric effector, A, to the activator site, j, induces a new conformation to the enzyme, one that has a greater affinity for the substrate. Binding of a negative allosteric effector to the inhibitor site, i, results in an enzyme conformation having a decreased affinity for substrate. (b) A model of a polymeric allosteric enzyme. Binding of the positive allosteric effector, A, at the j site causes an allosteric change in the conformation of the protomer to which the effector binds. This change in the conformation is transmitted to the second protomer through cooperative protomer–protomer interactions. The affinity for the substrate is increased in both protomers. A negative effector decreases the affinity for substrate of both protomers.

oligomeric enzyme consisting of two protomers is visualized. In both models heterotropic interactions can occur between the activator and substrate sites. In model *b*, homotropic interactions can occur between the activator sites or between the substrate sites.

Kinetics of Allosteric Enzymes

As a consequence of the interaction between the substrate site, the activator site, and the inhibitor site, a characteristic sigmoid or S-shaped curve, as shown in Figure 4.18*a* (curve *A*), is obtained in *S* vs v_0 plots of allosteric enzymes. Negative allosteric effectors move the curve toward higher substrate concentrations and enhance the sigmoidicity of the curve. If we employ $\frac{1}{2}V_{max}$ as a guideline, it can be seen from Figure 4.18 that a higher concentration of substrate would be required to achieve $\frac{1}{2}V_{max}$ in the presence of a negative effector (curve *C*) than is required in the absence of negative effector (curve *A*). In the presence of a positive modulator (curve *B*), $\frac{1}{2}V_{max}$ can be reached at a lower substrate concentration than is required in the absence of the positive modulator (curve *A*). Positive modulators shift the v_0 vs *S* plots toward the hyperbolic plots observed in Michaelis–Menten kinetics.

From the viewpoint of metabolic control, allosteric enzymes allow fine control of the activity of individual enzymes through small fluctuations in the level of substrate. Often the in vivo concentration of substrate corresponds with the sharply rising segment of the sigmoid *S* vs v_0 plot; consequently, large changes in enzyme activity are effected by small changes in substrate concentration. (See Figure 4.18.) It is also possible to "turn the enzyme off" with small amounts of a negative allosteric effector by its shifting the apparent K_m to higher values, values which are far above the in vivo level of substrate. Notice that at a given in vivo concentration of substrate the

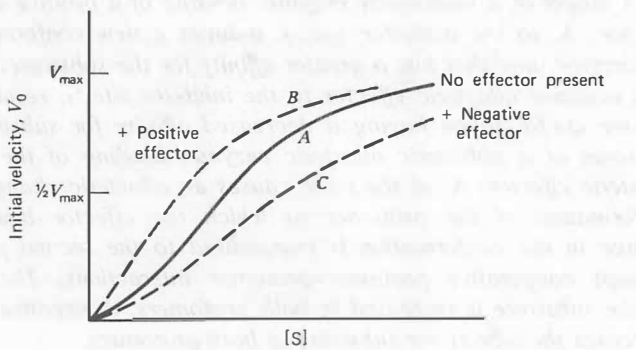


Figure 4.18
The kinetic profile of a K class allosteric enzyme.
The enzyme shows sigmoid S *vs* v_0 *plots. Negative effectors shift the curve to the right resulting in an increase in* K_m. *Positive effectors shift the curve to the left and effectively lower the apparent* K_m. V_{max} *is not changed.*

initial velocity, v_0, is decreased in the presence of a negative inhibitor (compare curves A and C).

Cooperativity

The realization that most allosteric enzymes are oligomeric and also have sigmoid S vs v_0 plots led to the concept of cooperativity to explain the interaction between ligand sites in oligomeric enzymes. Cooperativity is defined as the influence that the binding of ligand to one protomer has on the binding of ligand to a second protomer of an oligomeric protein. It should be emphasized that kinetic mechanisms other than cooperativity can also produce sigmoid v_0 vs S plots; consequently sigmoidicity is not diagnostic of cooperativity in a v_0 vs S plot. We have previously discussed cooperativity in terms of the binding of oxygen to hemoglobin and will now expand the concept.

The relationship between allosterism and cooperativity has been confused in many recent texts. The conformational changes occurring in a given protomer in response to ligand binding at an allosteric site is an allosteric effect. Cooperativity generally involves a change in conformation of an effector-activated protomer, which in turn transforms an adjacent protomer into a new conformation with an altered affinity for the effector ligand or for a second ligand. The conformation change may be induced by an allosteric effector or it may be induced by substrate, as it is in the case of hemoglobin. In hemoglobin the oxygen binding site on each protomer corresponds to the substrate site on an enzyme rather than to an allosteric site. Therefore, the oxygen-induced conformational change in the hemoglobin protomers is technically not an allosteric effect, although some authors identify it as such. It is a homotropic cooperative interaction. Those who consider the oxygen-induced changes in hemoglobin to be "allosteric" are using the term in a much broader sense than the original definition allows; however, "allosteric" is now used by many to describe any ligand-induced change in the tertiary structure of a protomer.

It should be emphasized that one can have an allosteric effect in the absence of any cooperativity. For example, in alcohol dehydrogenase, conformational changes can be demonstrated in each of the protomers upon the addition of positive allosteric effectors, but the active site of each protomer is completely independent of the other and there is no cooperativity between protomers, that is, induced conformational changes in one protomer are not transmitted to adjacent protomers.

In an attempt mathematically to describe experimentally observed ligand saturation curves, several models of cooperativity have been

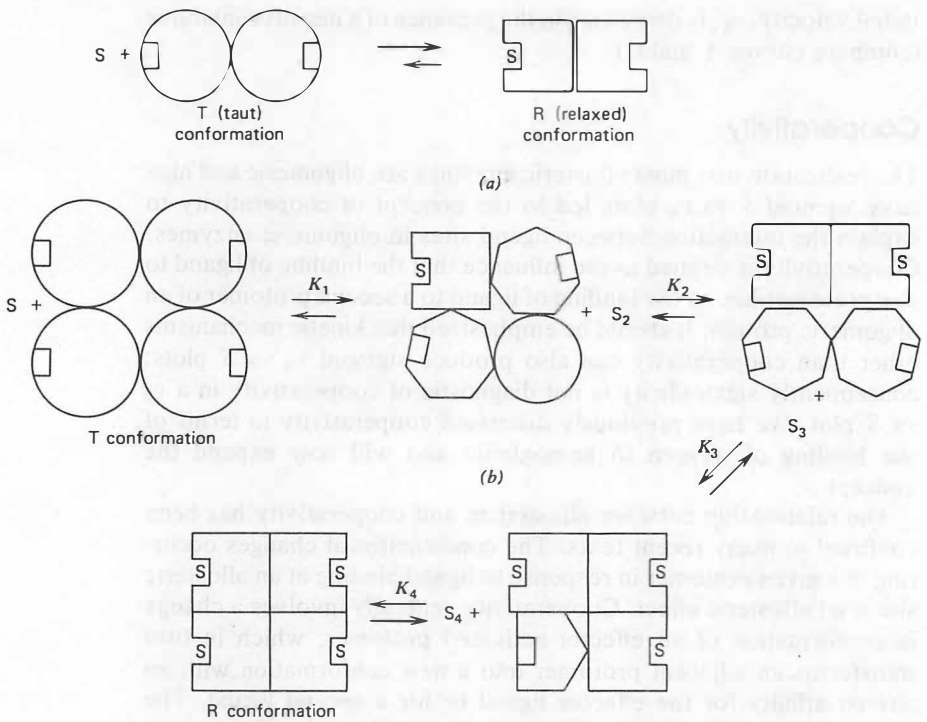


Figure 4.19
Models of cooperativity.
(a) The concerted model. The enzyme exists in only two states, the T, taut, and the R, relaxed, conformation. Substrates and activators have a greater affinity for the R state and inhibitors for the T state. Ligands shift the equilibrium between the T and R states. (b) The sequential induced fit model. Binding of a ligand to any one subunit induces a conformational change in that subunit. This conformational change is transmitted partially to adjoining subunits through subunit–subunit interaction. Thus the effect of the first ligand bound is transmitted cooperatively and sequentially to the other subunits (protomers) in the oligomer resulting in a sequential increase or decrease in ligand affinity of the other protomers. The cooperativity may be either positive or negative, depending upon the ligand.

proposed. The two most prominent models are the concerted and the sequential-induced fit.

Although the concerted model is rather restrictive, most of the nomenclature associated with allosterism and cooperativity arose from this model. The model proposes that the enzyme exists in only two states, the T (taut) and the R (relaxed). The T and R states are in

equilibrium. Activators and substrates favor the R state and shift the preexisting equilibrium toward the R state by the law of mass action. Inhibitors favor the T state. A conformational change in one protomer causes a corresponding change in all protomers. No hybrid states occur. The model is diagrammed in Figure 4.19*a*. Although the model accounts for the kinetic behavior of many enzymes, it cannot account for negative cooperativity.

The sequential-induced fit model proposes that ligand binding induces a conformational change in a protomer. A corresponding conformational change is then partially induced in an adjacent protomer contiguous with the protomer containing the bound ligand. The effect of ligand binding is sequentially transmitted through the oligomer, giving rise to an increasing or decreasing affinity for the ligand by contiguous protomers as suggested by the scheme in Figure 4.19*b*. In this model numerous hybrid states occur giving rise to the cooperativity and the sigmoid S vs v_0 plots. Both positive and negative cooperativity can be accommodated by the model. A positive modulator induces a conformation in the protomer, which has an increased affinity for the substrate. A negative modulator induces a different conformation into the protomer; one that has a decreased affinity for substrate. Both effects are cooperatively transmitted to adjacent protomers. For the V class enzymes the same reasoning applies, but the effect is on the catalytic event (k_3) rather than on K_m.

Regulatory Subunits

In the foregoing discussion the allosteric site has been considered to reside on the same protomer as the catalytic site. Furthermore, all protomers were considered to be identical. There are several very important enzymes in which a distinct regulatory protomer exists. These regulatory subunits have no catalytic function per se, but binding with the catalytic protomer modulates the activity of the catalytic subunit through an induced conformational change in the catalytic subunit.

Ribonucleotide reductase converts ribonucleotide diphosphates to deoxyribonucleotides for DNA biosynthesis (page 656). The active site is generated at the interface of two different subunits, B_1 and B_2. Both B_1 and B_2 are catalytically inactive by themselves. The allosteric sites reside on B_1; however, B_2 does not contain the active site in and of itself. The active site is generated at the interface between polymerized B_1 and B_2, as suggested in Figure 4.20. ATP is a positive effector, which binds to the *l* allosteric site, and the end product of the enzyme, dATP, is a negative effector, which binds at the R site. This particular arrangement of subunits to form the catalytic site is not common.

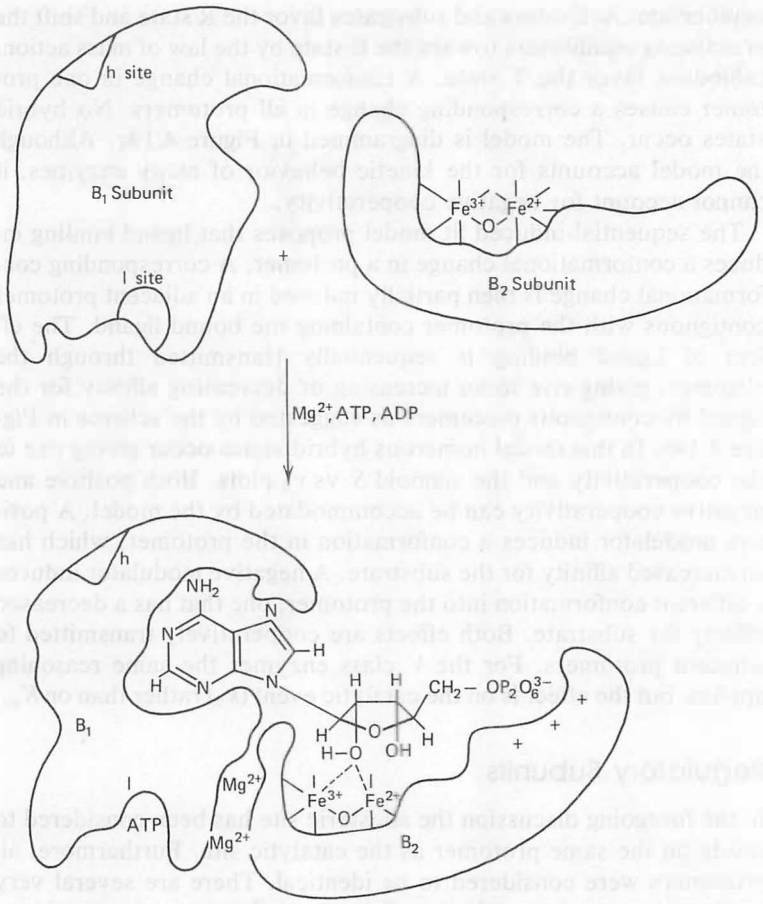


Figure 4.20
Model of an allosteric enzyme with a separate regulatory subunit.
The enzyme is ribonucleotide reductase. B_1 is the regulatory subunit. ATP binds at the "I site" and is a positive effector. B_2 is the catalytic subunit only when combined with the B_1 subunit. B_2 contains the iron cluster, which is involved in reduction of the 2'-hydroxyl of the ribose of the ribonucleotide diphosphates, which in this example is ADP. dATP is a negative modifier that competes at the "I" site for ATP. Other modifiers such as dTTP bind at the "h" site and are either positive or negative modulators depending on the substrate as shown in Figure 4-39.

Completely separate regulatory subunits are observed in aspartate transcarbamylase. This enzyme is involved in pyrimidine biosynthesis and catalyzes the transfer of the carbamoyl group from carbamoyl phosphate to the α-amino group of aspartate:

Carbamoyl phosphate + aspartate → carbamoylaspartate + P_i

The enzyme is a complex consisting of two catalytic (C) and three regulatory (R) subunits. The end product of the pyrimidine pathway, cytidine triphosphate (CTP), binds to the R subunits as a negative effector, whereas ATP functions as a positive effector.

Two other very important enzymes, adenyl cyclase and protein kinase, have regulatory subunits. They are schematically depicted in

Figures 4.40 and 4.42. These two particular enzymes are discussed in detail on page 204.

4.7 ENZYME SPECIFICITY: THE ACTIVE SITE

Enzymes are the most specific catalysts known, both from the viewpoint of the substrate as well as the type of reaction performed on the substrate. Specificity inherently resides in the substrate binding site, which lies on the enzyme surface. The tertiary structure of the enzyme is folded in such a way as to create a region that has the correct molecular dimensions, the appropriate topology, and the optimal alignment of counterionic groups and hydrophobic regions to accommodate a specific substrate. The tolerances in the active site are so small that usually only one isomer of a diastereomeric pair will bind. For example, D-amino acid oxidase will bind only D-amino acids and not L-amino acids. Some enzymes show absolute specificity for substrate. Others have broader specificity and will accept several different analogs of a specific substrate. For example, hexokinase catalyzes the phosphorylation of glucose, mannose, fructose, glucosamine, and 2-deoxyglucose, but not all at the same rate. Glucokinase, on the other hand, is specific for glucose.

The specificity of the reaction catalyzed rests in the active site and the particular arrangement of amino acids that participate in the bond-making and bond-breaking phase of catalysis. The mechanism of catalysis is discussed in Section 4.8.

Complementarity of Substrate and Enzyme

Various models have been proposed to explain the substrate specificity of enzymes. The first proposal was the "lock-and-key" model (Figure 4.21), in which a negative impression of the substrate exists

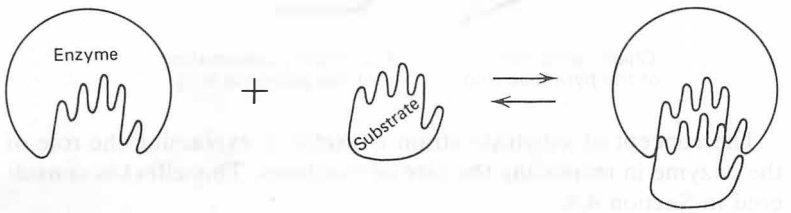


Figure 4.21
Lock and key model of the enzyme binding site.
The enzyme contains a negative impression of the molecular features of the substrate, thus allowing specificity of the enzyme for a particular substrate.

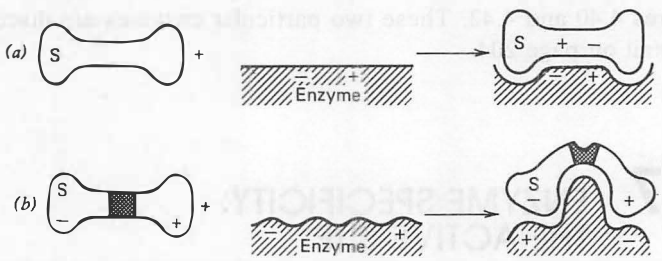
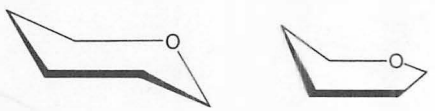


Figure 4.22
Models for induced fit and substrate strain. *(a) Approach of substrate to the enzyme induces the formation of the active site. (b) Substrate strain, induced by substrate binding to the enzyme, contorts normal bond angles and "activates" the substrate.* Reprinted, with permission, from D. Koshland and P. Boyer, *Annu. Rev. of Biochem.*, 37:374, 1968. Copyrighted by Annual Reviews, Inc.

on the enzyme surface. The substrate fits to this binding site just as a key goes into the proper lock or a hand into the proper sized glove. This model gives a rigid picture of the enzyme and cannot account for the effects of allosteric ligands.

A more flexible model of the binding site is the *induced fit* model. In this model, the binding site and certainly the active site are not fully formed. The essential elements of the binding site are present to the extent that the correct substrate can position itself properly in the nascent binding site. The initial interaction of substrate with the enzyme induces a conformational change in the tertiary structure of the enzyme, resulting in the formation of the strong binding site and the repositioning of the appropriate catalytic amino acids to form the active site. There is excellent x-ray evidence for the correctness of this model in the enzyme carboxypeptidase A. A diagram of the induced fit model is shown in Figure 4.22*a*.

The contemporary concept of *induced fit* combined with substrate strain accounts for more of the experimental observations concerning enzyme action than do other models. In this model (Figure 4.22*b*), the substrate is "strained" toward product formation as a result of an induced conformational transition of the enzyme. A good example of enzyme-induced substrate strain is observed in lysozyme (Figure 4.23) where the conformation of the sugar residue "D" at which bond breaking occurs is strained from the stable chair to the unstable half-chair conformation upon binding to the enzyme:



Chair conformation of the pyranose ring

Half-chair conformation of the pyranose ring

The concept of substrate strain is useful in explaining the role of the enzyme in increasing the rate of reactions. This effect is considered in Section 4.8.

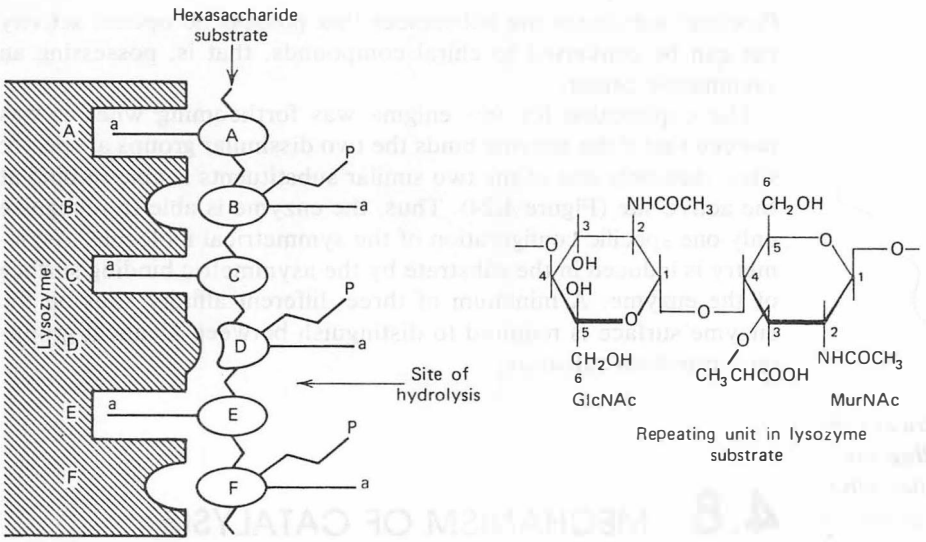
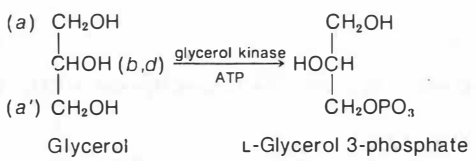


Figure 4.23
Hexasaccharide binding at the active site of lysozyme.
In the model substrate pictured, the ovals represent glucose. Ring D is strained by the enzyme to the half-chair conformation and hydrolysis occurs between the D and E rings. Six subsites on the enzyme bind substrate. Alternate sites are specific for acetamido groups but are unable to accept the lactyl (P) side chains, which occur on the N*-acetylmuramic acid residues. Thus the substrate can bind to the enzyme in only one orientation.*

Reprinted, with permission, from T. Imoto, L. N. Johnson, A. C. T. North, et al. *The Enzymes*, 3rd, P. Boyer, ed. 7:713, 1972. Academic Press.

Asymmetry of the Binding Site

Not only are enzymes able to distinguish isomers, but they are also able to make a distinction between two equivalent atoms in a symmetrical molecule. For example, the enzyme glycerol kinase is able to distinguish between configurations of H and OH on carbon-2 in the symmetric substrate glycerol, so that the only asymmetric product formed is L-glycerol 3-phosphate. These prochiral substrates have two identical substituents and two additional but dissimilar groups on the same carbon ($C_{aa'bd}$).



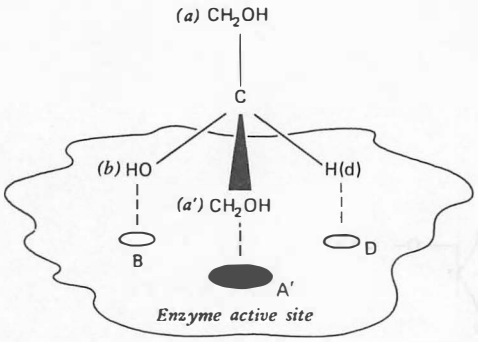


Figure 4.24
Three-point attachment of a symmetrical substrate to an asymmetric enzyme binding site. *Glycerol kinase by virtue of dissimilar binding sites for the —H and —OH group of glycerol binds only the a' hydroxymethyl group to the active site. Thus, only one stereoisomer results from the kinase reaction, the* L*-glycerol 3-phosphate.*

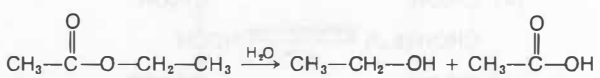
Prochiral substrates are substances that possess no optical activity but can be converted to chiral compounds, that is, possessing an asymmetric center.

The explanation for this enigma was forthcoming when it was noticed that if the enzyme binds the two dissimilar groups at specific sites, then only one of the two similar substituents is able to bind at the active site (Figure 4.24). Thus, the enzyme is able to recognize only one specific configuration of the symmetrical molecule. Asymmetry is induced in the substrate by the asymmetric binding surface of the enzyme. A minimum of three different affinity sites on the enzyme surface is required to distinguish between identical groups on a prochiral substrate.

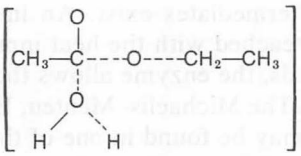
4.8 MECHANISM OF CATALYSIS

All chemical reactions have a potential energy barrier that must be overcome before reactants can be converted to products. In the gas phase the reactant molecules can be given enough kinetic energy by heating so that collisions result in product formation. The same is true with solutions. However, a well-controlled body temperature of 37°C does not allow temperature to be increased to accelerate the reaction, and 37°C is not warm enough to provide the reaction rates required for fast-moving species of animals. Enzymes employ other means of overcoming the barrier to reaction, and these will be discussed after some useful definitions are covered.

A comparison of the energy diagrams for catalyzed and noncatalyzed reactions is shown in Figure 4.25. The energy barrier represented by the uncatalyzed curve in Figure 4.25 is a measure of the *activation* energy, E_a. The reaction coordinate is simply the pathway in terms of bond stretching between reactants and products. At the apex of the energy barrier is the activated complex known as the *transition state,* Ts. The transition state represents the reactants in their activated state. In this state reactants are in an intermediate stage along the reaction pathway and cannot be identified as starting material or products. For example, in the hydrolysis of ethyl acetate:



the Ts might look like



The transition state complex can break down to products or go back to reactants. The Ts is not an intermediate and cannot be isolated!

Notice that in the case of the enzyme-catalyzed reaction (Figure 4.25) the energy of the reactants and products is no different than in the uncatalyzed reaction. Enzymes do not change the thermodynamics of the system but they do change the pathway for reaching the final state.

As noted on the energy diagram, there may be several plateaus or valleys on the energy contour for an enzyme reaction. At these

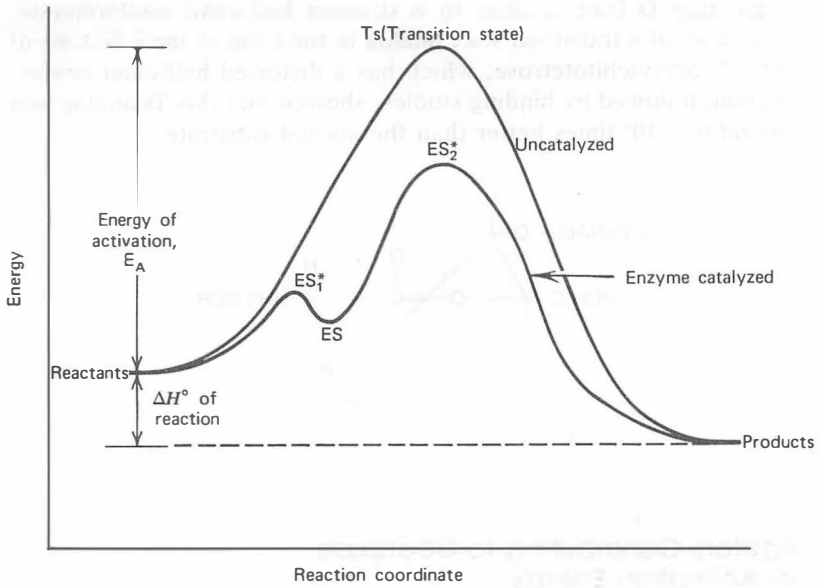
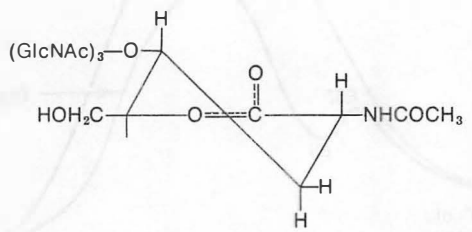


Figure 4.25
Energy diagrams for catalyzed vs noncatalyzed reactions.
The overall energy difference between reactants and products is the same in catalyzed and noncatalyzed reactions. The enzyme-catalyzed reaction proceeds at a faster rate because the energy of activation is lowered. The standard heat of formation of the products is $\Delta H°$.

Modified, with permission, from *Review of Biochemistry*, N. H. Sloan and J. L. York, page 46. 1969 Macmillan Company.

points metastable intermediates exist. An important point is that each valley may be reached with the heat input available in a 37°C system. In other words, the enzyme allows the energy barrier to be scaled in increments. The Michaelis–Menten, ES, complex is not the transition state, but may be found in one of the valleys. This is the case because in the ES complex, the substrates are properly oriented and the substrate may be "strained," and therefore the bonds to be broken lie further along the reaction coordinate.

If our concepts of the transition state are correct, one would expect that compounds designed to closely resemble the transition state would bind much more tightly to the enzyme than does the natural substrate. This has proven to be the case. In such substrate analogs one finds affinities 10^2 to 10^5 times greater than substrate affinity. These compounds are called *transition state* analogs and are potent enzyme inhibitors. Previously, lysozyme was discussed in terms of substrate strain, and mention was made of the conversion of sugar ring D from a chair to a strained half-chair conformation. Synthesis of a transition state analog in the form of the δ-lactone of tetra-*N*-acetylchitotetrose, which has a distorted half-chair conformation, followed by binding studies, showed that this Ts analog was bound 6×10^3 times better than the normal substrate.



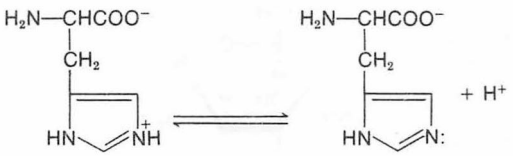
Factors Contributing to Decrease in Activation Energy

Enzymes are able to enhance the rates of reaction by a factor of 10^9 to 10^{12} times that of the noncatalyzed reaction. Most of this rate enhancement can be accounted for by four processes: acid–base catalysis, substrate strain, covalent catalysis, and entropy effects.

Acid–Base Catalysis

Specific acids and bases are H^+ and OH^-, respectively. Free protons and hydroxide ions are not encountered in most enzyme reactions and then only in some metal-dependent enzymes (page 160).

General acids and bases are important in enzymology. A *general* acid or base is any substance that is weakly ionizable. In the physiological pH range, the protonated form of histidine is the most important general acid and its conjugate base an important general base. Other acids are the thiol —SH, tyrosine —OH and the ϵ-amino group of lysine. Other bases are carboxylic acid anions and the conjugate bases of the general acids.



Histidine as a general acid

Histidine as a general base

Consideration of the mutarotation of glucose illustrates the principle of acid–base catalysis. If one dissolves pure α-D-glucose in water, over several days an equilibrium mixture of 31% α-D-glucose

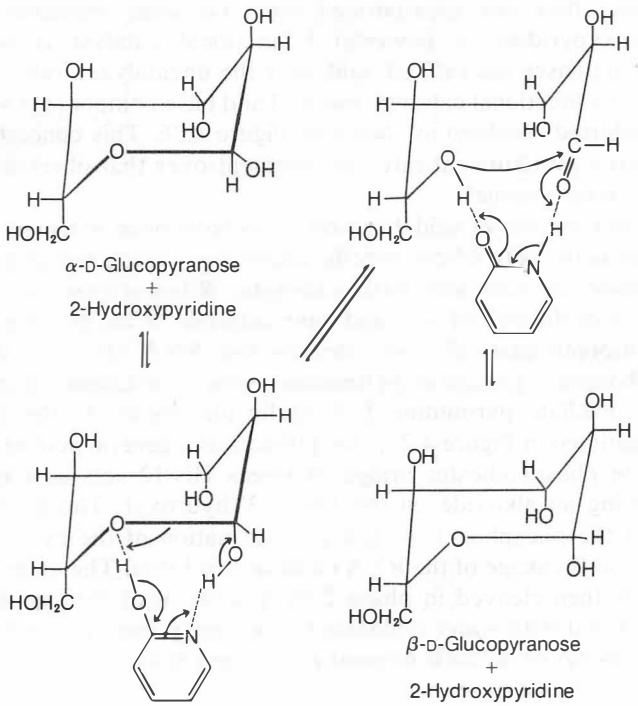
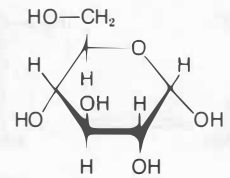
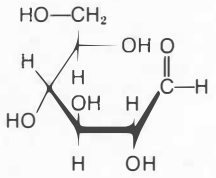


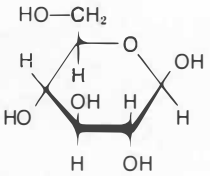
Figure 4.26
Example of concerted acid–base catalysis.
Mutarotation of glucose is enhanced 10^5-fold by 2-hydroxypyridine, which contains both an acidic and basic functional group. The phenolic hydroxyl protonates the pyranose ring oxygen simultaneous with the generation of intramolecular nucleophile by the pyridine base. The latter opens the pyranose ring yielding the free aldehyde. Reclosure of the ring is not a stereospecific reaction; hence the product is either the α- or β-glucopyranose.



α-D-Glucopyranose



D-Glucose
(open chain aldehyde
form)



β-D-Glucopyranose

and 64% β-D-glucose forms. The conversion of the α anomer to the β anomer proceeds through intermediate formation of the open-chain aldehyde. Upon reforming the pyranose ring, there is no stereospecificity in the orientation of the resulting hydroxyl on carbon-1 (the anomeric carbon), and the ring can reclose with the hydroxyl in either the α or β orientation.

General acids increase the rate of mutarotation by protonating the pyranose oxygen which labilizes the C-—O bond yielding the aldehyde. The rate of mutarotation is increased 1,000-fold over the uncatalyzed reaction. General bases increase the rate another 1,000-fold over the acid rate by generating an alkoxide ion from the hydroxyl on C_1 which quickly opens the pyranose ring. If the general acid and base are incorporated into the same molecule as in 2-hydroxypyridine, a powerful bifunctional catalyst is formed, which increases the rate 10^5-fold over the uncatalyzed rate.

In the bifunctional catalyst, the acid and base components work in a "concerted" fashion as shown in Figure 4.26. This concerted action gives an additional rate enhancement over that observed with either process alone.

These concepts of acid–base catalysis have been extended to the enzyme active site where specific amino acid functional groups are implicated as acid and base catalysts. Ribonuclease is a good example of the role of acid and base catalysis at the enzyme active site. Ribonuclease (RNase) cleaves the RNA chain at the 3′-phosphodiester linkage of pyrimidines with an obligatory formation of intermediate pyrimidine 2′,3′-cyclic phosphate. In the mechanism outlined in Figure 4.27, His-119 acts as a general acid to protonate the phosphodiester bridge, whereas His-12 acts as a base in generating an alkoxide on the ribose-3′-hydroxyl. The latter then attacks the phosphorus, resulting in formation of the cyclic phosphate and breakage of the RNA chain at this locus. The cyclic phosphate is then cleaved in phase 2 by a reversal of the reactions in phase 1, but with water replacing the leaving group. The active site histidines revert to their original protonated state.

Substrate Strain

Our previous discussion of this topic related to induced fit of enzymes to substrate. It is also possible that binding of substrate to a preformed site on the enzyme induces strain into the substrate. Irrespective of the mechanism of strain induction, the important point is that the energy level of the substrate is raised, and the substrate is propelled toward the bonding found in the transition state.

A combination of substrate strain and acid–base catalysis is observed in the mechanism of lysozyme action (Figure 4.28). Ring D of the hexasaccharide substrate upon binding to the enzyme is strained

to the half-chair conformation. General acid catalysis by active site glutamic acid promotes the unstable half-chair into the transition state. The carbonium ion formed in the transition state is stabilized by a negatively charged aspartate. Breakage of the glycosidic linkage between rings D and E relieves the strained transition state by allowing rings D and E to return to the stable chair conformation.

Covalent Catalysis

In covalent catalysis, the attack of a nucleophilic (negatively charged) or electrophilic (positively charged) group in the enzyme

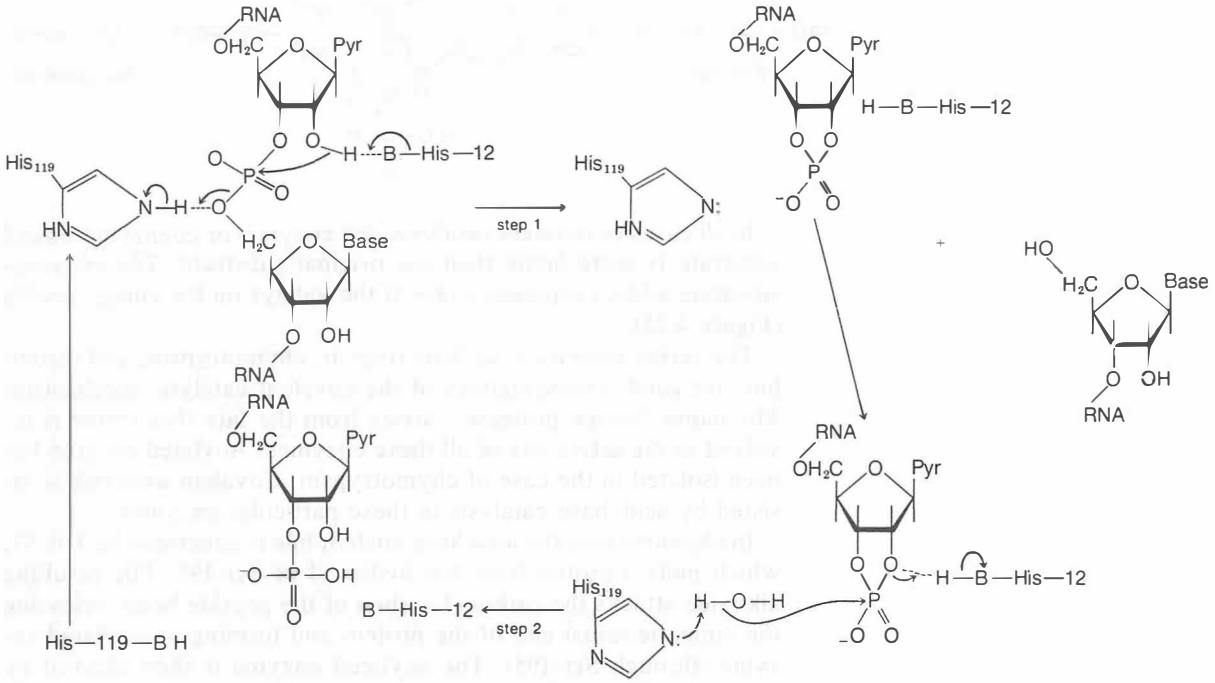


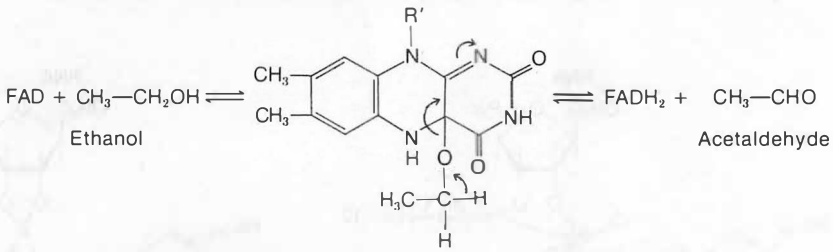
Figure 4.27

Role of acid and base catalysis in the active site of ribonuclease.

RNase cleaves the phosphodiester bond in pyrimidine loci in ribonucleic acid. Histidine-119 and histidine-12 function as acid and base catalysts respectively, in enhancing the formation of an intermediate 2',3'-cyclic phosphate and release of a shorter fragment of RNA. These same histidines then play a reverse role in the hydrolysis of the cyclic phosphate with release of the other fragment of RNA that ends in a pyrimidine 3'-phosphate.

Modified, with permission, from *Chemistry and Control of Enzyme Reactions*, K. G. Scrimgeour, page 193. 1977 Academic Press.

active site upon the substrate results in covalent binding of the substrate to the enzyme as an intermediate in the reaction sequence. Also, enzyme-bound coenzymes often form covalent bonds with the substrate. For example, in the transaminases, the amino acid substrate forms a Schiff's base with enzyme-bound pyridoxal phosphate (page 546). Evidence is now accumulating that some oxidoreductases, utilizing FAD as coenzyme, form intermediate covalent adducts of substrate and FAD. For example, in the oxidation of alcohols the scheme shown here has been proposed.



In all cases of covalent catalysis, the enzyme- or coenzyme-bound substrate is more labile than the original substrate. The enzyme–substrate adduct represents one of the valleys on the energy profile (Figure 4.25).

The serine proteases, such as trypsin, chymotrypsin, and thrombin, are good representatives of the covalent catalytic mechanism. The name "serine protease" arises from the fact that serine is involved in the active site of all these enzymes. Acylated enzyme has been isolated in the case of chymotrypsin. Covalent catalysis is assisted by acid–base catalysis in these particular enzymes.

In chymotrypsin the attacking nucleophile is generated by His-57, which pulls a proton from the hydroxyl of Ser-195. The resulting alkoxide attacks the carbonyl carbon of the peptide bond, releasing the amino-terminal end of the protein and forming an acylated enzyme (through Ser-195). The acylated enzyme is then cleaved by reversal of the reaction sequence, but with water as the nucleophile rather than Ser-195. This mechanism is outlined in Figure 4.29. It was formerly believed that aspartic acid 102 increased the basicity of histidine 57 through a "charge relay," but this has now been shown not to be the case.

Entropy Effect

Entropy is a thermodynamic term, S, which defines the extent of disorder in a system. At equilibrium, the entropy is maximal. For example, in solution two reactants A—⟨ and B—◇ exist in many different orientations. The chances of A—⟨ and B—◇ coming to-

gether with the correct geometric orientation and with enough energy to react is small at 37°C and in dilute solution. However, if an enzyme with two high-affinity binding sites for A—◁ and B—◇ is introduced into the dilute solution of these reactants, as suggested in Figure 4.30, A—◁ and B—◇ will be bound to the enzyme in the correct orientation for the reaction to occur. They will be bound with the correct stoichiometry, and the effective concentration of the reactants will be increased on the enzyme surface—all of which will contribute to an increased rate of reaction.

Once correctly positioned on the enzyme surface, and as a result of binding, the substrates may be "strained" toward the transition

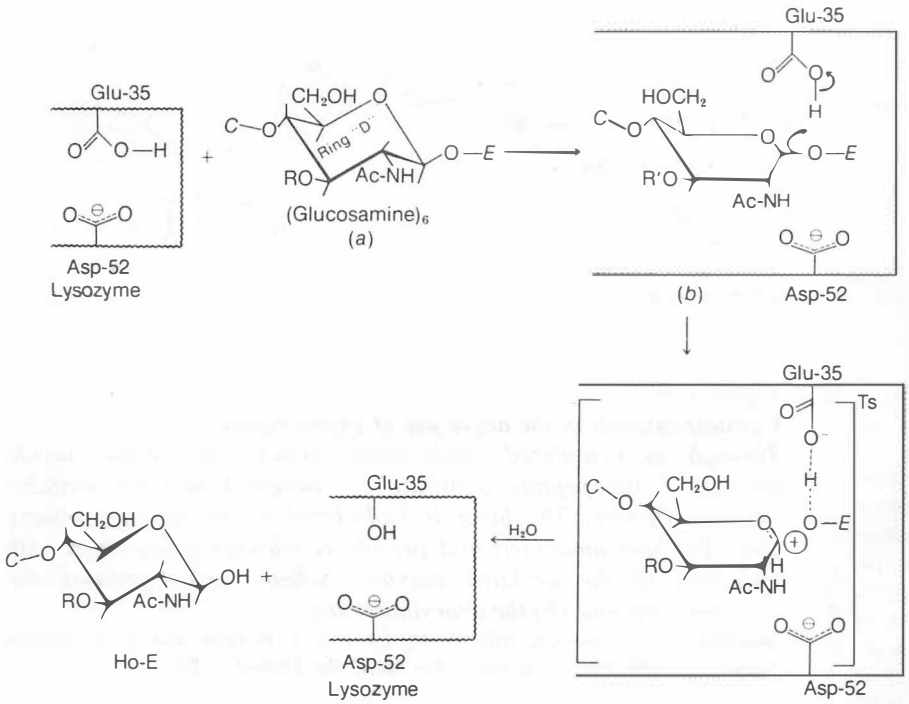


Figure 4.28
A mechanism for lysozyme action–substrate strain.
The binding of the stable chair (a) conformation of the substrate to the enzyme generates the strained half-chair conformation (b) in the ES complex. In the transition state, acid-catalyzed hydrolysis of the glycosidic linkage by an active site glutamic acid residue generates a carbonium ion on the D ring, which relieves the strain generated in the initial ES complex and results in collapse of the transition state to products.

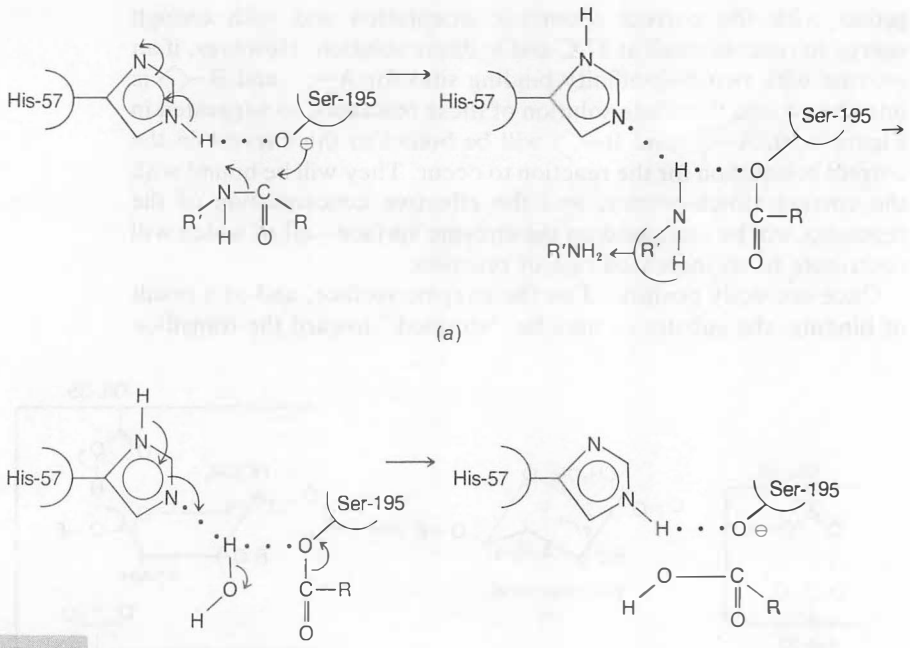


Figure 4.29
Covalent catalysis in the active site of chymotrypsin.
Through acid-catalyzed nucleophilic attack, the stable amide linkage of the peptide substrate is converted into an unstable acylated enzyme. The latter is hydrolyzed in the rate-determining step. The new amino-terminal peptide is released concomitant with formation of the acylated enzyme. Scheme (a) represents the acylation step and (b) the deacylation step.
Modified, with permission, from D. M. Blow, J. J. Birktoft, and B. S. Hartley, *Nature*, 221:339, 1969. Copyright, 1969 Macmillan Journal LTD).

state. At this point the substrates have been "set up" for acid–base and/or covalent catalysis. Thus, the proper orientation and the nearness of the substrate with respect to the catalytic groups, which has been dubbed the "proximity effect," contributes 10^3–10^4-fold to the rate enhancement observed with enzymes. It has been estimated that the decrease in entropy contributes a factor of 10^3 to the rate enhancement.

Environmental Effects on Catalysis

A number of external parameters, including pH, temperature, and salt concentration, affect the enzyme activity. These effects are

CLIN. CORR. 4.4 THERMAL LABILITY OF G6PD RESULTS IN A HEMOLYTIC ANEMIA

Glucose 6-phosphate dehydrogenase is an important enzyme in the red cell in terms of the maintenance of the red cell membrane (chapter 21). A deficiency or inactivity of this enzyme leads to a hemolytic anemia. In other cases, a variant enzyme is present that normally has sufficient activity to maintain the membrane, but under conditions of oxidative stress fails. A reasonably frequent mutation of this enzyme leads to a protein with normal kinetic constants but a decreased thermal stability. This condition is especially critical to the red cell, since it is devoid of protein-synthesizing capacity and cannot renew enzymes as they denature. The end result is a greatly decreased lifetime for those red cells which have an unstable G6P dehydrogenase. These red cells are also susceptible to drug-induced hemolysis.

probably not important in vivo, under normal conditions, but are very important in setting up enzyme assays in vitro to measure enzyme activity in a patient's plasma or tissue sample.

Temperature Dependence

Plots of velocity vs temperature for most enzymes reveal a bell-shaped curve with an optimum between 40 and 45°C for mammalian enzymes, as indicated in Figure 4.31. Above this temperature, heat denaturation of the enzyme occurs, while at lower temperatures there is not enough heat to overcome the energy barrier, even for the catalyzed reaction. Between 0 and 40°C, most enzymes show a two-fold increase in activity for every 10° rise in temperature. Under conditions of hypothermia most enzyme reactions are depressed, which accounts for the decreased oxygen demand of living organisms at low temperature. Mutation of an enzyme to a thermolabile form can have serious consequences as discussed in Clin. Corr. 4.4.

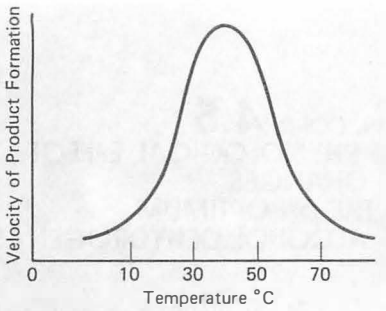


Figure 4.31
Temperature dependence of a typical mammalian enzyme.
To the left of the optimum the rate is low because the environmental temperature is too low to provide enough kinetic energy to overcome the energy of activation. To the right of the optimum, the enzyme is inactivated by heat denaturation.

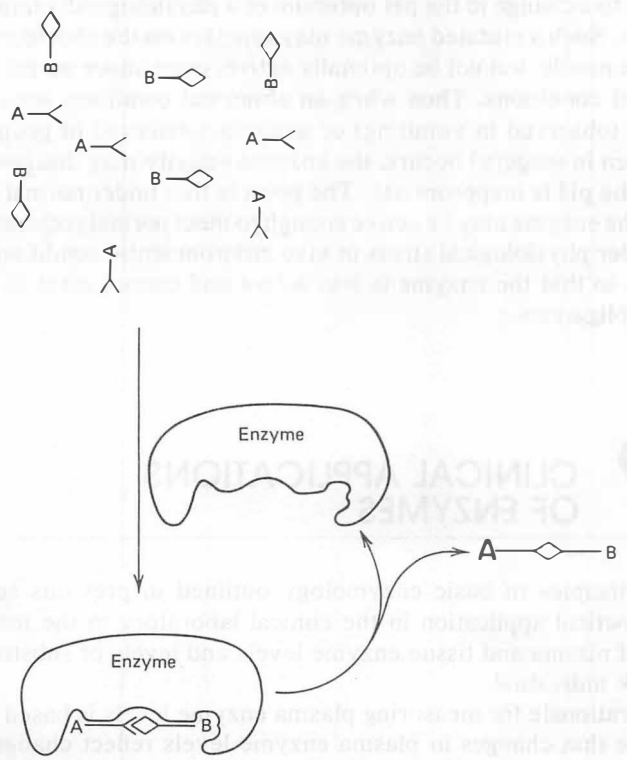


Figure 4.30
Diagrammatic representation of the entropy effect.
Substrates in dilute solution are concentrated and oriented on the enzyme surface so as to enhance the rate of reaction.

CLIN. CORR. 4.5 THE PHYSIOLOGICAL EFFECT OF CHANGES IN THE pH OPTIMUM OF ALCOHOL DEHYDROGENASE

The unusual sensitivity of orientals to alcoholic beverages may have a biochemical basis. In the case of Japanese and Chinese, much less ethanol is required to produce the vasodilation that results in facial flushing and tachycardia than is required to achieve the same effect in Europeans. These physiological effects arise from the acetaldehyde generated by liver alcohol dehydrogenase

$$CH_3CH_2OH + NAD^+ \rightleftharpoons CH_3CHO + H^+ + NADH$$

Alcohol dehydrogenase has three separate genetic loci for the production of three polypeptide chains-α, β, and γ. Now two alleles have been found in the loci responsible for the β chain such that two β chains are produced, β_1 and β_2. The correlation of ethnic background and percentage of the β_2 chain in alcohol dehydrogenase is shown in the table.

Nationality	*Percent of Total β Chain as β_2*
White American	5–10
Swiss	20
Japanese	85

The alcohol dehydrogenase containing a high percentage of the β_2 chain has a shift in pH optimum from the normal pH of 10 to 8.5. The end result of the subtle change

pH Effects

Nearly all enzymes show a bell-shaped pH–velocity profile, but the maximum (pH optimum) varies greatly with the particular enzyme. Alkaline and acid phosphatase are both found in man, but their pH optima are greatly different, as shown in Figure 4.32. Obviously, neither functions at maximal activity at the pH of blood.

The bell-shaped curve and its position on the X axis are dependent upon the particular ionized state of the substrate that will be optimally bound to the enzyme. This in turn is related to the ionization of specific amino acids that constitute the substrate binding site. In addition, those amino acids that are involved in catalyzing the reaction must be in the correct state ionically to be functional in the catalytic event. For example, if aspartic acid is involved in catalyzing the reaction, the pH optimum may be in the region of 4.5 where the α-carboxyl of aspartate ionizes, whereas if the ε-amino of lysine is the catalytic group, the pH optimum may be around pH 9.5, the pK_a of the ε-amino group. Studies of the pH dependence of enzymes are useful for suggesting which amino acid(s) may be operative in the catalytic event in the active site.

Clin. Corr. 4.5 points out the physiological effect of a mutation leading to a change in the pH optimum of a physiologically important enzyme. Such a mutated enzyme may function on the shoulder of the pH-rate profile, but not be optimally active, even under normal physiological conditions. Then when an abnormal condition such as alkalosis (observed in vomiting) or acidosis (observed in pneumonia and often in surgery) occurs, the enzyme activity may disappear because the pH is inappropriate. The point is that under normal conditions, the enzyme may be active enough to meet normal requirements, but under physiological stress in vivo environmental conditions may change so that the enzyme is less active and cannot meet its metabolic obligations.

4.9 CLINICAL APPLICATIONS OF ENZYMES

The principles of basic enzymology outlined in previous sections find practical application in the clinical laboratory in the measurement of plasma and tissue enzyme levels and levels of substrates in the sick individual.

The rationale for measuring plasma enzyme levels is based on the premise that changes in plasma enzyme levels reflect changes that

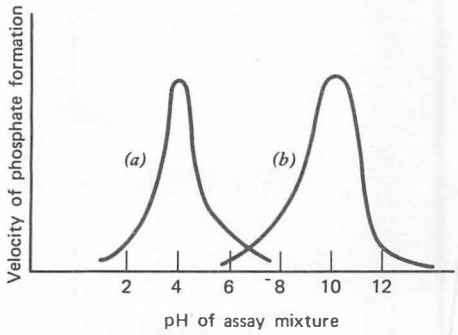


Figure 4.32
The pH dependence of (a) acid and (b) alkaline phosphatase reactions. *In each case the optimum represents the ideal ionic state for binding of enzyme and substrate and the correct ionic state for the amino acids involved in the catalytic event.*

in pH dependence is a marked increase in activity of alcohol dehydrogenase in Japanese because of the shift in pH dependence toward more physiological pH. More acetaldehyde is produced quickly after ethanol ingestion, resulting in a greater physiological response.

have occurred in a specific tissue or organ. Plasma enzymes are of two types: one type is present in the highest concentration and has a functional role; the other type is normally present at very low levels and plays no functional role in the plasma. The former type includes the enzymes associated with blood coagulation (thrombin), fibrin dissolution (plasmin), and processing of chylomicra (lipoprotein lipase).

In terms of our interest in disease of tissues and organs other than blood, the nonplasma specific enzymes are most important. Normally, the plasma levels of these enzymes are low to absent. An insult in the form of any disease process may cause changes in cell membrane permeability or increased cell death, resulting in release of intracellular enzymes into the plasma. In cases of permeability change, those enzymes of lower molecular weight will appear in the plasma first. The greater the concentration gradient between intra- and extracellular levels, the more rapidly the enzyme diffuses out. Cytoplasmic enzymes will appear in the plasma before mitochondrial enzymes, and of course the greater the quantity of tissue damaged, the greater the increase in the plasma level. The nonplasma specific enzymes will be cleared from the plasma at varying rates, which depend upon the stability of the enzyme and its susceptibility to the reticuloendothelial system.

In the diagnosis of specific organ involvement in a disease process it would be ideal if enzymes unique to each organ could be identified; however, this is unlikely, since the metabolism of various organs is not unique. Alcohol dehydrogenase of the liver and acid phosphatase

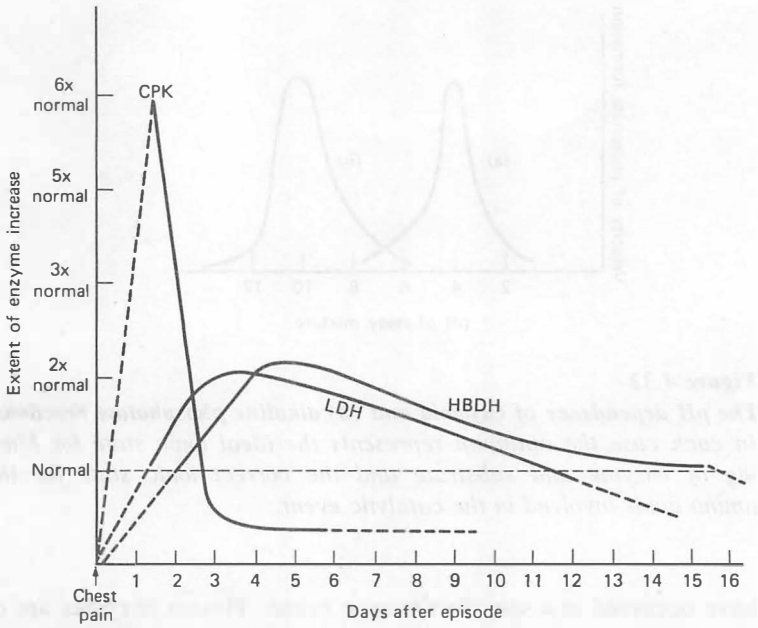


Figure 4.33
Kinetics of release of cardiac enzymes into serum following a myocardial infarction. *Creatine phosphokinase, CPK; lactic dehydrogenase, LDH; serum glutamic-oxalacetic transaminase, SGOT; β-hydroxybutyric dehydrogenase, HBDH. Such kinetic profiles allow one to determine where the patient is with respect to the infarct and recovery. Note: CPK rises sharply but briefly; HBDH rises slowly but persists.*

Reprinted, with permission, from *Diagnostic Enzymes*, E. L. Coodley, page 61. 1970 Lea and Febiger.

of the prostate are useful for specific identification of disease in these organs. Other than these two examples there are few additional enzymes that are organ-specific. The situation is saved because the ratio of various enzymes does vary from tissue to tissue. This fact, combined with a study of the kinetics of appearance and disappearance of particular enzymes from the plasma, allows a diagnosis of specific organ involvement to be made. Figure 4.33 illustrates the time dependence of the plasma levels of enzymes released from the myocardium following a heart attack. Such profiles allow one to establish when the attack occurred and whether treatment is effective.

Clin. Corr. 4.6 demonstrates how diagnosis of a specific enzyme defect led to a rational clinical treatment that restored the patient to health.

Studies of the kinetics of appearance and disappearance of plasma enzymes are predicated on a valid enzyme assay. The establishment of a good assay is based on maintaining good temperature and pH control, as well as saturating levels of all substrates, cosubstrates, and cofactors. In order to accomplish the latter, the K_m must be known for those particular conditions of pH, ionic strength, and so on, that are to be used in the assay. You will recall that K_m is the substrate concentration at half-maximal velocity (V_{max}). To be assured the system is saturated, the substrate concentration is gener-

CLIN. CORR. 4.6 IDENTIFICATION AND TREATMENT OF AN ENZYME DEFICIENCY

Enzyme deficiencies usually lead to increased accumulation of specific intermediary metabolites in plasma and hence in urine. The recognition of the intermediates that are accumulating is useful in pinpointing possible enzyme defects. After the enzyme deficiency is established, metabolites that normally occur in the pathway but are distal to the block may be supplied exogenously in order to overcome the metabolic effects of the enzyme deficiency. For example, in hereditary orotic aciduria there is a double enzyme

ally increased 5- to 10-fold over the K_m. The importance of saturation of the enzyme with substrate is that only under these conditions is the reaction zero order. This fact is emphasized in Figure 4.34. Only under zero-order conditions are changes in velocity proportional to enzyme concentration alone. Under first-order conditions in substrate, the velocity is dependent upon both the substrate and the enzyme concentration.

Clin. Corr. 4.7 demonstrates the importance of determining if the assay conditions accurately reflect the amount of enzyme actually present. Clinical laboratory assay conditions are optimized routinely for the properties of the normal enzyme and may not reflect levels of mutated enzyme. The pH dependence and/or the K_m for substrate and cofactors may drastically change in a mutated enzyme.

Under optimal conditions a valid enzyme assay will reflect a linear dependence of velocity and amount of enzyme. This can be tested by determining if the velocity of the reaction doubles when the plasma sample size is doubled, while keeping the total volume of the assay constant, as demonstrated in Figure 4.35.

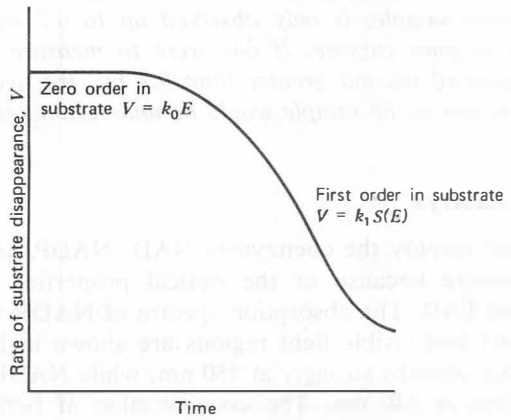
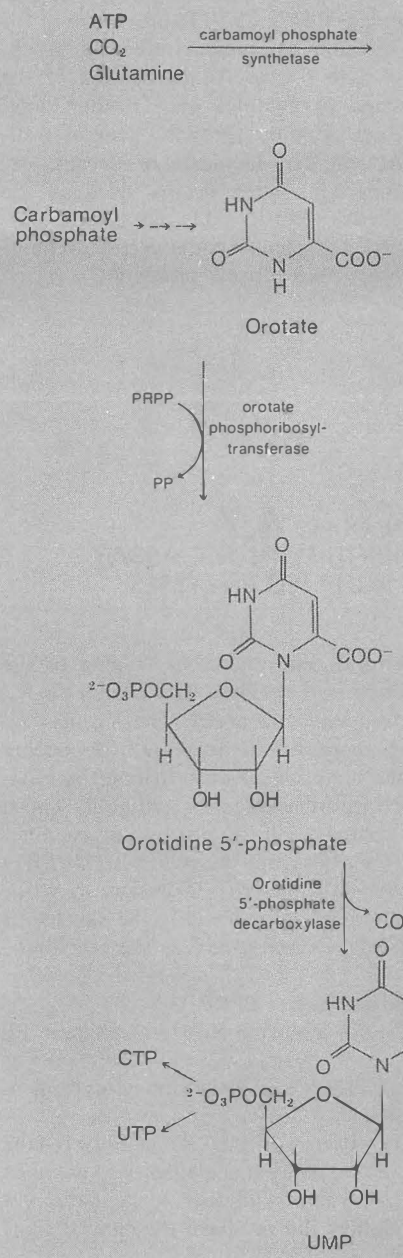


Figure 4.34
Relation of substrate concentration to order of the reaction.
When the enzyme is completely saturated, the kinetics are zero order with respect to substrate and are first order in enzyme, that is, the rate depends only on enzyme concentration. When the substrate level falls below saturating levels, the kinetics are first order in both substrate and enzyme and are therefore second order, that is, the observed rate is dependent upon both enzyme and substrate.

Reprinted, with permission, from *Review of Biochemistry*, N. H. Sloan and J. L. York, page 48. 1969 Macmillan Company.

deficiency in the pyrimidine biosynthetic pathway leading to accumulation of orotic acid.



Both orotate phosphoribosyltransferase and orotidine 5'-phosphate decarboxylase are deficient. Since the pyrimidine nucleotides, CTP and TTP, are required for cell division, the patients are pale, weak, and fail to thrive. Administration of the missing pyrimidines as uridine and cytidine promotes growth, general well-being, and also decreases orotic acid excretion. The latter occurs because the supplied TTP and CTP repress the carbamoyl phosphate synthase, the committed step, by feedback inhibition.

CLIN. CORR. 4.7 AMBIGUITY IN THE ASSAY OF MUTATED ENZYMES

Structural gene mutations leading to the production of enzymes with changes in K_m are frequently observed. The K_m may be either increased or decreased, depending on the mutation. A case in point is a patient with hyperuricemia and gout, whose red blood cell hypoxanthine-guanine-phosphoribosyltransferase (HGPRT) showed little activity in assays in vitro. This enzyme is involved in the salvage of purine bases and catalyzes the reaction

Hypoxanthine + PPRP ⟶ inosine monophosphate + PP

where PPRP = phosphoribosyl pyrophosphate.

The absence of HGPRT activity results in a severe neurological disorder known as Lesch-Nyhan syndrome (page 645), yet this patient did not have the clinical signs

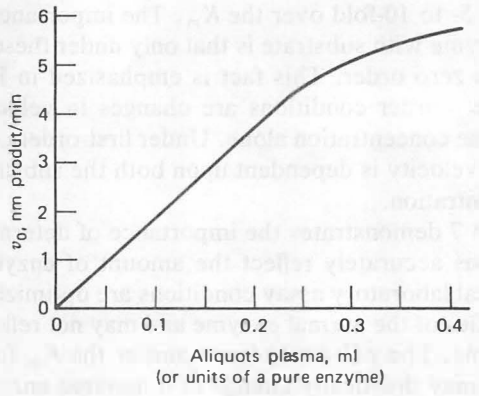


Figure 4.35
Assessing the validity of an enzyme assay.
The line shows what is to be expected for any reaction where the concentration of substrate is held constant and the aliquots of enzyme increased. In this particular example linearity between the initial velocity observed and the amount of enzyme, whether pure or in a plasma sample, is only observed up to 0.2 ml of plasma or 0.2 units of pure enzyme. If one were to measure the velocity with an aliquot of plasma greater than 0.2 ml, the actual amount of enzyme present in the sample would be underestimated.

Coupled Assays

Enzymes that employ the coenzymes NAD, NADP, and FAD are easy to measure because of the optical properties of NADH, NADPH, and FAD. The absorption spectra of NADH and FAD in the ultraviolet and visible light regions are shown in Figure 4.36. Oxidized FAD absorbs strongly at 450 nm, while NADH has maximal absorption at 340 nm. The concentration of both FAD and NADH is related to their absorption of light at the respective absorption maximum by the Beer–Lambert relation

$$A = \varepsilon \cdot c \cdot l$$

where l is the pathlength of the spectrometer cell in centimeters (usually 1 cm), ε is the absorbance of a molar solution of the substance being measured at a specific wavelength of light, A is the absorbance read off the spectrometer, and c is the concentration. Absorbance is the log of transmittance (I_0/I). ε is a constant which varies from substance to substance; its value can be found in a handbook of biochemistry. In an optically clear solution, the concen-

tration c can be found after a determination of the absorbance A is made.

Many enzymes do not employ either NAD or FAD but do generate products that can be utilized by a NAD- or FAD-linked enzyme. For example, glucokinase catalyzes the reaction

$$\text{Glucose} + \text{ATP} \rightleftharpoons \text{glucose 6 phosphate} + \text{ADP}$$

Both ADP and G6P are difficult to measure directly; however, the enzyme glucose 6-phosphate dehydrogenase catalyzes the reaction,

$$\text{G6P} + \text{NADP} \rightleftharpoons \text{6-phosphogluconolactone} + \text{NADPH} + \text{H}^+$$

Thus by adding an excess of the enzyme G6P dehydrogenase and NADP to the assay mixture, the velocity of production of G6P by glucokinase is proportional to the rate of reduction of NADP, which can be measured directly in the spectrophotometer.

Isoenzymes: Clinical Application

Isozymes (or *isoenzymes*) are enzymes that catalyze the same reaction but migrate differently on electrophoresis. Their physical properties may also be different, but not necessarily. The most common mechanism for the formation of isozymes involves the arrangement of subunits arising from two different genetic loci in different combinations to form the active polymeric enzyme. The isozymes that

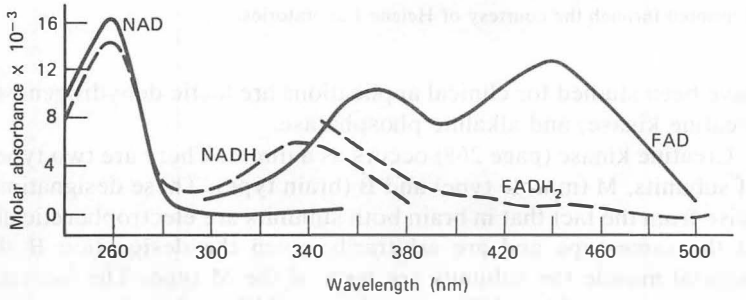


Figure 4.36
Absorption spectra of niacin and flavin coenzymes.
The reduced form of NAD (NADH) absorb strongly at 340 nm. The oxidized form of flavin coenzymes absorb strongly at 450 nm. Thus, one can follow the rate of reduction of NAD by observing the increase in the absorbance at 340 nm and the formation of $FADH_2$ by following the decrease in absorbance at 450 nm.

of this disorder. Furthermore, immunological testing with a specific antibody to the enzyme revealed as much cross-reacting material in the patient's red blood cells as in normal controls. The conclusion is that the enzyme is being produced but is inactive in the assay in vitro. Additional experimentation revealed that by increasing the substrate concentration in the assay, full activity was measurable in the patient's red cell hemolysates. This anomaly is explained as a mutation in the substrate binding site of HGPRT, leading to an increased K_m. Neither the substrate concentration in the assay nor in the red blood cells was high enough to bind to the enzyme. This case reinforces the point that an accurate enzyme determination is dependent upon zero-order kinetics, that is, the enzyme is saturated with substrate.

Why doesn't the patient have the Lesch–Nyhan syndrome? Most likely because the cells in his nervous tissue are not affected by the mutation. This is possible since the red cell originates from a different stem cell line than the cells of the nervous system.

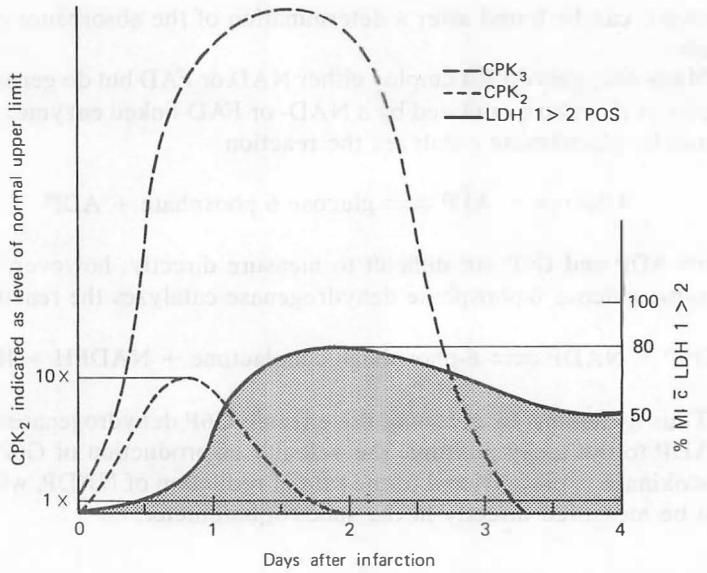


Figure 4.37
Characteristic changes in serum CPK and LDH isozymes following a myocardial infarction.
CPK_2 (MB) isozyme increases to a maximum within 1 day of the infarction. CPK_3 lags behind CPK_2 by about 1 day. The total LDH level increases more slowly. The increase of LDH_1 and LDH_2 within 12–24 h coupled with an increase in CPK_2 is diagnostic of myocardial infarction.
Reprinted through the courtesy of Helena Laboratories.

have been studied for clinical applications are lactic dehydrogenase, creatine kinase, and alkaline phosphatase.

Creatine kinase (page 268) occurs as a dimer. There are two types of subunits, M (muscle type) and B (brain type). These designations arise from the fact that in brain both subunits are electrophoretically of the same type and are arbitrarily given the designation B. In skeletal muscle the subunits are both of the M type. The isozyme containing both M and B type subunits (MB) is found only in the myocardium. Other tissues contain variable amounts of the MM and BB isozymes. The isozymes are numbered beginning with the species migrating the fastest to the anode thus, CPK_1 (BB), CPK_2 (MB), and CPK_3 (MM).

Lactic dehydrogenase is a tetrameric enzyme, but only two distinct subunits have been found: those designated H for heart (myocardium) and M for muscle. These two subunits are combined

in five different ways. The lactic dehydrogenase isozymes, subunit compositions, and major location are as follows:

Type	*Composition*	*Location*
LDH_1	HHHH	Myocardium and RBC
LDH_2	HHHM	Myocardium and RBC
LDH_3	HHMM	Brain and kidney
LDH_4	HMMM	
LDH_5	MMMM	Liver and skeletal muscle

As an illustration of how measurement of amounts of isozymes and kinetic analysis of plasma enzyme levels are useful in medicine, levels of some CPK and LDH isozymes are plotted in Figure 4.37 as a function of time after infarction. After damage to heart tissue the cellular breakup releases CPK_2 into the blood within the first 6–18 h after an infarct, but LDH release lags behind the appearance of CPK in the plasma by 1 to 2 days. Normally the LDH_2 isozyme is higher in concentration than LDH_1; however, in the case of infarction the amount of LDH_1 becomes greater than LDH_2, as shown in Figure 4.37 at about the time CPK_2 levels are back to baseline (48–60 h). Figure 4.38 shows the fluctuations of all five isozymes after an infarct. The flip of LDH_2 and LDH_1 can be seen in the 24-h tracing. The LDH isozyme "switch" coupled with an increased CPK_2 is diagnostic of MI in virtually 100% of the cases. Increased levels of LDH_5 are an indicator of liver congestion. Thus secondary complications of heart failure can be monitored.

Formerly, plasma levels of the two transaminases SGOT (serum glutamic-oxalacetic transaminase) and SGPT (serum glutamic-pyruvic transaminase) were followed; however, these enzymes have much less specificity and predictive accuracy in diagnosing MI and liver disease. The rationale for assaying these two enzymes is that liver and heart contain high levels of both enzymes, but liver contains more GPT than GOT and the reverse is true in heart.

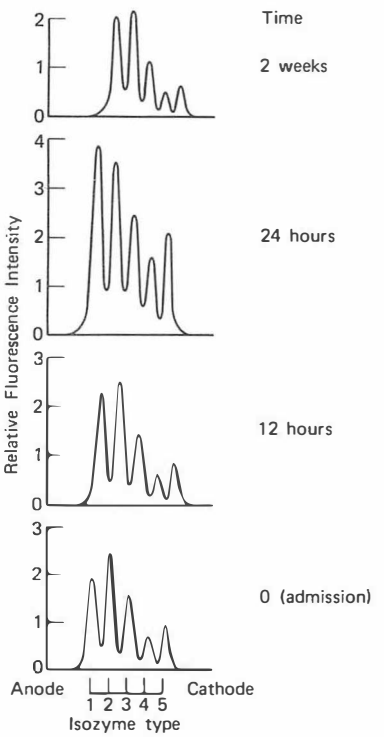


Figure 4.38
Tracings of densitometer scans of LDH isozymes at time intervals following a myocardial infarction.
As can be seen total LDH increases and LDH_1 becomes greater than LDH_2 between 12 and 24 h. Increases in LDH_5 is diagnostic of a secondary congestive liver involvement. After electrophoresis on agarose gels the LDH activity is assayed by measuring the fluorescence of the NADH formed in the LDH catalyzed reaction.
Courtesy of Dr. A. T. Gajda, Clinical Laboratories, The University of Arkansas for Medical Science.

Enzymes as Therapeutic Agents

In a few cases enzymes have been used as drugs in the therapy of specific medical problems. Streptokinase is an enzyme mixture prepared from a streptococcus. It is useful in clearing blood clots that occur in the lower extremities. Streptokinase activates the fibrinolytic proenzyme plasminogen that is normally present in plasma. The activated enzyme is plasmin. Plasmin is a serine protease like trypsin that attacks fibrin, cleaving it into several soluble components.

Asparaginase therapy is used for some types of adult leukemia. Tumor cells have a nutritional requirement for asparagine and must scavenge it from the host's plasma. By administering asparaginase i.v., the host's plasma level of asparagine is markedly depressed, which results in depressing the viability of the tumor.

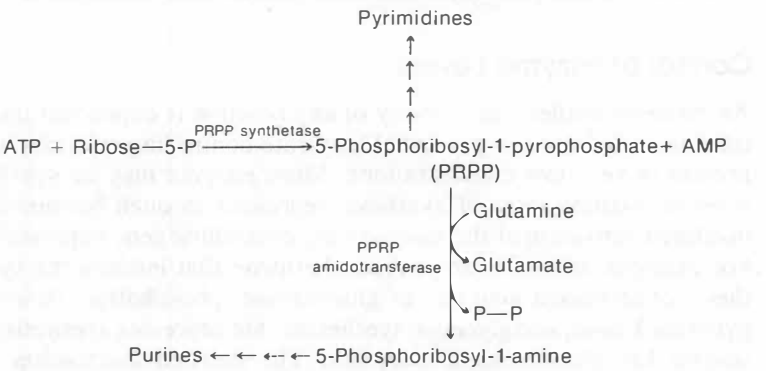
Most enzymes do not have a long half-life in blood; consequently, unreasonably large amounts of enzyme are required to keep the therapeutic level up. Work is now in progress to enhance enzyme stability by coupling enzymes to solid matrices and implanting these materials in areas that are well perfused. In the future, enzyme replacement in individuals that are genetically deficient in a particular enzyme may be feasible.

4.10 REGULATION OF ENZYME ACTIVITY

Our discussion up to this point has centered upon the chemical and physical characterization of individual enzymes, but physiologically we must be concerned with the integration of many enzymes into a metabolic pathway and the interrelationship of the products of one pathway with the metabolic activity of other pathways. For example, dietary glucose can be either converted to glycogen, fat, some nonessential amino acids or oxidized to carbon dioxide. In each case, glucose is converted to a different end product through a specific metabolic pathway involving several enzymes each of which is unique to the type of reaction catalyzed. After eating, there is an abundance of glucose in the system, but it is not diverted in equal amounts to each of the end products mentioned. Rather, there are very tightly controlled homeostatic mechanisms, which work to maintain a constant blood glucose level, utilize the glucose needed for energy production, maintain the glycogen stores, and, if any excess glucose remains, convert it to fat. The point is that all metabolic

pathways are not operating at maximum capacity at all times. In fact many pathways may be shut down during certain phases in the life cycle of a cell. If this were not the case, wild uncontrolled and uneconomical growth of the cell would occur.

Control of metabolic regulation of a pathway occurs through modulation of the enzymatic activity of one or more key enzymes in the pathway. Although the overall catalytic efficiency of a metabolic pathway is dependent upon the activity of all the individual enzymes in the pathway, the pathway can be controlled by one *rate-limiting* enzyme in the pathway. Usually this rate-controlling enzyme is the first enzyme that can be identified as unique to that particular pathway. The chemical reaction that is unique to a metabolic pathway is referred to as the *committed step.* For example, in the de novo synthesis of purines, the committed step is the reaction catalyzed by the PRPP amidotransferase, which in this case is also the rate-controlling enzyme. The rate-limiting enzyme is not necessarily the enzyme associated with the committed step. The substrate of the amidotransferase, PRPP, is also used as substrate by the pyrimidine biosynthetic pathway; hence the enzyme PRPP synthase, which produces PRPP, does not catalyze the committed step in the biosynthesis of purines because it occurs before the branch point in the two pathways. These relationships in the purine and pyrimidine pathways can be visualized as shown here.



The activity of the enzyme associated with the committed step or with the rate-limiting enzyme can be regulated in a number of ways. First, the absolute amount of the enzyme can be regulated either through substrate or hormonal stimulation of the de novo synthesis of more enzyme. Hormones can also suppress the de novo synthesis of enzyme. Second, the activity of the enzyme can be modulated by activators, inhibitors and by covalent modification through mechanisms previously discussed. Finally, the activity of a pathway can be

regulated by partitioning the pathway from its initial substrate and by controlling access of the substrate to the enzymes of the pathway. This is referred to as *compartmentation.* We will now consider each of these general mechanisms of control in more detail.

Compartmentation

Generally anabolic and catabolic pathways are segregated into different organelles in order to maximize the cellular economy. There would be no point to the oxidation of fatty acids occurring at the same time and in the same compartment as biosynthesis of fatty acids. If such occurred, a futile cycle would exist. By maintaining fatty acid biosynthesis in the cytoplasm and oxidation in the mitochondria, control can be exerted by regulating transport of common intermediates across the mitochondrial membrane. For example, coenzyme A derivatives of fatty acids cannot diffuse across the mitochondrial membrane but are transported by a specific transport system. If the metabolic situation requires fatty acid biosynthesis rather than fatty acid oxidation, there could be hormonal or other control over the mitochondrial membrane fatty acid transport system such that it is depressed during fatty acid biosynthesis, but activated when fatty acid oxidation is required for cellular energy.

Table 4.2 contains a compilation of some of the important enzymes, metabolic pathways, and their intracellular distribution.

Control of Enzyme Levels

As indicated earlier, the velocity of any reaction is dependent upon the amount of enzyme present. Many rate-controlling enzymes are present in very low concentrations. More enzyme may be synthesized or existing rates of synthesis repressed through hormonally instituted activation of the mechanisms controlling gene expression. For example, insulin is an anabolic hormone that induces the synthesis of increased amounts of glucokinase, phosphofructokinase, pyruvate kinase, and glycogen synthetase, but represses synthesis of several key gluconeogenic enzymes. The detailed mechanism of these effects are not known in mammalian systems; however, the general concepts of the regulation of eucarotic gene expression are discussed in Chapter 20.

In some instances substrate can repress the synthesis of enzyme. For example, glucose represses the de novo synthesis of pyruvate carboxykinase. This enzyme is the rate-limiting enzyme in the conversion of pyruvate to glucose. In other words, if there is plenty of glucose available there is no point in synthesizing glucose at the expense of amino acids which are the alternative source of pyruvate;

Table 4.2 Intracellular Location of Major Enzymes and Metabolic Pathways[a]

Cytoplasm	Glycolysis; hexose monophosphate pathway; glycogenesis and glycogenolysis; fatty acid synthesis; purine and pyrimidine catabolism; peptidases; aminotransferases; amino acyl synthetases
Mitochondria	Krebs citric acid cycle; fatty acid oxidation; amino acid oxidation; fatty acid elongation; urea synthesis; electron transport and coupled oxidative phosphorylation
Lysosomes	Lysozyme; acid phosphatase; hydrolases, including proteases, nucleases, glycosidases, arylsulfatases, lipases, phospholipases and phosphatases
Endoplasmic reticulum (microsomes)	NADH and NADPH cytochrome c reductases; cytochrome b_5 and cytochrome P_{450} related mixed function oxidases; glucose 6-phosphatase; nucleoside diphosphatase, esterase, β-glucuronidase, and glucuronyltransferase; protein synthetic pathways; phosphoglyceride and triacylglycerol synthesis; steroid synthesis and reduction
Golgi	Galactosyl- and glucosyltransferase; chondroitin sulfotransferase; 5′-nucleotidase; NADH-cytochrome c reductase; glucose 6-phosphatase
Peroxisomes	Urate oxidase; D-amino acid oxidase; α-hydroxy acid oxidase; catalase
Nucleus	DNA and RNA biosynthetic pathways

[a] NADH-cytochrome b_5 reductase has been found in endoplasmic reticulum, Golgi, outer mitochondrial membrane, and in the nuclear envelope. Several of the enzymes noted in the table are common to one or more of the membranous organelles.

consequently, this pathway is repressed by the effect of its end product, glucose, on the synthesis of the carboxykinase enzyme. This effect of glucose may be mediated via insulin as discussed on page 775 and is not direct feedback inhibition.

Many rate-controlling enzymes have relative short half-lives; for example, that of pyruvate carboxykinase is 5 h. Teleologically this is reasonable because it provides a mechanism for effecting much larger fluctuations in the activity of a pathway than would be possible by inhibition or activation of existing levels of enzyme.

Regulation by Modulation of Activity

Regulation at the gene level is long-term. Short-term regulation occurs through modification of the activity of existing levels of enzyme

by means of mechanisms we have previously discussed and will now integrate into a generalized statement of control.

During various phases of the cell cycle, specific metabolic pathways are turned on or off, depending on the special requirements of a given phase of the cell cycle. For example, there is no point to continued production of deoxyribonucleotides at all times during a cell's life, but only during replicative phases; consequently, during nonreplicative phases, the concentration of deoxyribonucleotide builds up to such an extent that the diphosphoribonucleotide reductase is inhibited by the end products of the pathway. This type of control is referred to as *feedback inhibition*. The inhibition may take the form of competitive inhibition, allosteric inhibition, or any other mechanism of inhibition discussed in Section 4.5. In any case, the apparent K_m may be raised above the in vivo levels of substrate, and the reaction ceases or decreases in velocity. In addition to feedback within the pathway, feedback on other pathways also occurs. This is referred to as cross-regulation. An example of cross-regulation is observed in the production of deoxyribonucleotides where dGTP not only inhibits its own synthesis but also that of dTTP, as indicated in Figure 4.39. Such cross-regulation is essential for DNA synthesis, since all four bases must be present at the same time and in approximately equal concentrations. Activation of the system occurs by ATP-induced enhancement of the CDP and UDP reductase activities by positive allosteric mechanisms. The end product of the UDP reductase and associated reactions (page 657) is dTTP, which activates the GDP reductase, leading to increased dGTP which in turn activates the ADP reductase activity. All these reductase activities are found in one complex enzyme, but the principles of cross-regulation cited can be applied to interpathway regulation.

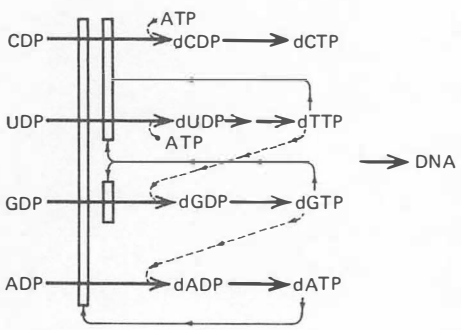


Figure 4.39
Cross-regulation in the synthesis of deoxynucleoside triphosphates.
Open bars indicate inhibition. Broken lines indicate activation. The enzyme affected is ribonucleotide reductase.

Reprinted, with permission, from L. Thelander and P. Reichard, *Annual Review of Biochemistry*, 48:153, 1979. Copyright, Annual Reviews Inc.

Regulation by Covalent Modification

The first example of this mechanism of regulation was glycogen phosphorylase, in which the intraconvertible a (active) and b (inactive) forms were recognized to be phosphorylated and dephosphorylated enzymes, respectively.

Other examples of reversible covalent modification include acetylation–deacetylation, adenylation–deadenylation, uridylylation–deuridylylation, and methylation–demethylation.

The phosphorylation–dephosphorylation scheme is most common and will be considered in detail. There are four different modes of phosphorylation based upon the cofactor requirement. These are the cAMP dependent, the non-cAMP-dependent, the calcium-dependent, and the double-stranded RNA-dependent protein kinases. The cAMP dependent phosphorylation can be considered as

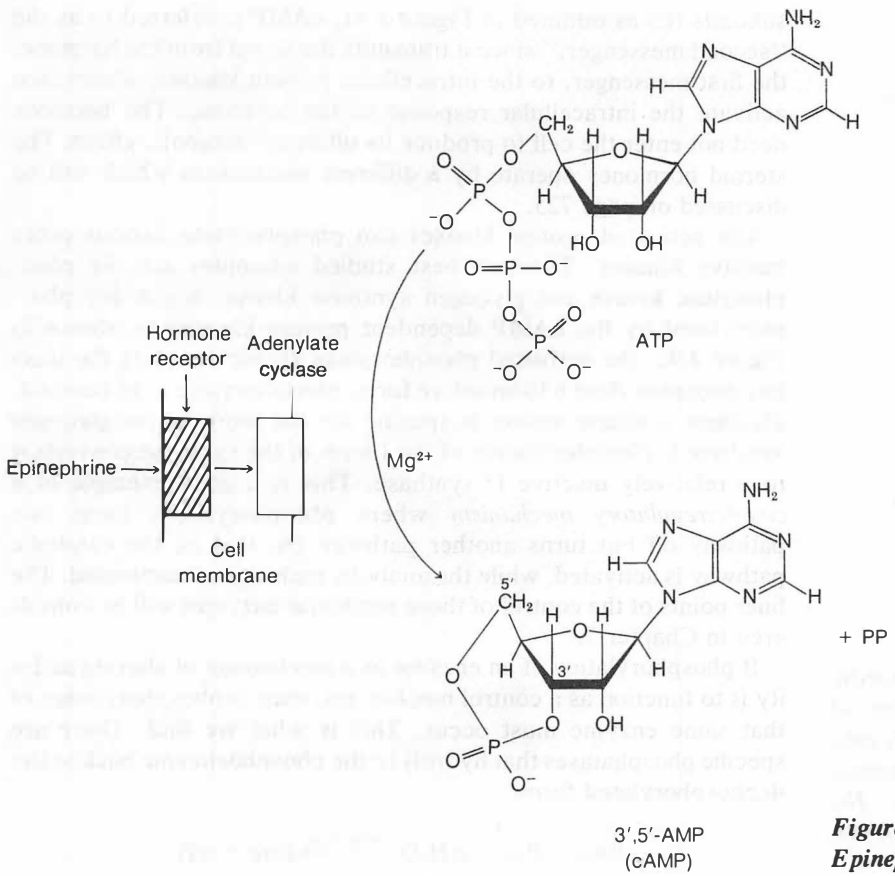


Figure 4.40
Epinephrine stimulation of adenyl cyclase.

characteristic of the other types of covalent modification. Details of the mechanisms will be covered in the chapters on metabolism and regulation.

cAMP-Dependent Phosphorylation

The phosphorylation of an enzyme occurs as the end result of a cascade of reactions initiated by the binding of a hormone such as epinephrine to a specific extracellular membrane receptor on the target tissue. Such binding activates the adenylate cyclase on the intracellular plasma membrane through an induced conformation change. Adenylate cyclase catalyzes the cyclization of ATP to cAMP, which then allosterically activates several protein kinases (Figure 4.40).

The cAMP activates the protein kinase by combining with its regulatory (R) subunits, resulting in the release of active catalytic

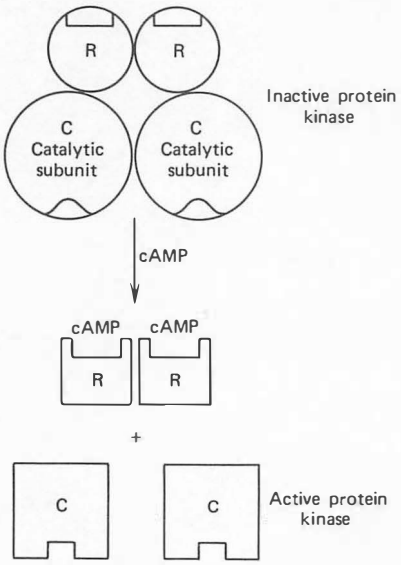


Figure 4.41
cAMP activation of protein kinase.
The enzyme is tetrameric with two catalytic and two regulatory subunits. Binding of cAMP to the regulatory subunits (R) results in dissociation of the complex and activation of the catalytic subunits (C). The active protein kinase is rather nonspecific with respect to the protein substrate.

subunits (C) as outlined in Figure 4.41. cAMP is referred to as the "second messenger," since it transmits the signal from the hormone, the first messenger, to the intracellular protein kinases, which then activate the intracellular response to the hormone. The hormone need not enter the cell to produce its ultimate metabolic effect. The steroid hormones operate by a different mechanism which will be discussed on page 725.

The activated protein kinases can phosphorylate various other inactive kinases. The two best studied examples are the phosphorylase kinase and glycogen synthase kinase, which are phosphorylated by the cAMP dependent protein kinase. As shown in Figure 4.42, the activated phosphorylase kinase converts the inactive phosphorylase b to an active form, phosphorylase a. In contrast, glycogen synthase kinase is specific for the more active glycogen synthase I. Phosphorylation of the I form of the synthase converts it to a relatively inactive D synthase. This is a good example of a *counterregulatory mechanism* where phosphorylation turns one pathway off but turns another pathway on, that is, the catabolic pathway is activated, while the anabolic pathway is inactivated. The finer points of the control of these particular enzymes will be considered in Chapter 7.

If phosphorylation of an enzyme as a mechanism of altering activity is to function as a control mechanism, then dephosphorylation of that same enzyme must occur. This is what we find. There are specific phosphatases that hydrolyze the phosphoenzyme back to the dephosphorylated forms:

$$\text{Enz—}(P_i)_n + nH_2O \xrightarrow{\text{phosphatase}} \text{Enz} + nP_i$$

In most cases the phosphate is esterified to the enzyme via a serine or threonine hydroxyl, although in some enzymes a phosphohistidine occurs.

For phosphorylation–dephosphorylation to function in regulation, appropriate external signals should bring about changes in the relative concentrations of phosphorylated and dephosphorylated enzyme. Regulation could occur through control of the protein kinase step and/or through control of the phosphatase step. In fact regulation occurs in both steps. The protein kinase is sensitive to cAMP levels which in turn are sensitive to various hormone levels. The regulation of the phosphatase seems to be through substrate-directed effects rather than specific effectors like cAMP. In this case the metabolic regulator combines with the phosphoprotein substrate, rendering it a "good" or "poor" substrate for the phosphatase. For example, the phosphorylase a phosphatase reaction is inhibited by G1P and AMP but stimulated by caffeine, glucose, and glucose 6-

phosphate; that is, the configuration of the phosphorylase a in the presence of glucose and glucose 6-phosphate renders it a better substrate for the phosphatase than does AMP and G1P. These activators are allosteric effectors that change the conformation of the phosphorylase a in such a way that the phosphatase can attack the phosphate on the phoshorylase. These effects have important physiological implications (page 395).

Interaction between the phosphatase activity and the phosphorylation of the regulatory subunits of phosphorylase kinase has been observed. The β regulatory subunit is phosphorylated first, but is sensitive to dephosphorylation by the phosphatases. Phosphorylation of the α subunit does not increase the catalytic activity of the γ subunit per se, but does decrease the rate of dephosphorylation of the β subunit; hence phosphorylation of both α and β regulatory subunits enhances the activity of the enzyme. This may be a general mechanism of control for those enzymes containing two regulatory subunits that undergo phosphorylation.

In addition to specific phosphatases like the phosphorylase a phosphatase just described, there is a nonspecific phosphatase that is composed of regulatory and catalytic subunits. Heat-stable inhibitors of the enzyme have been isolated from muscle but are uncharacterized. The mechanism of control of this protein phosphatase is unclear.

Control of cAMP Levels

For effective regulation of any system, there must be an "off" as well as an "on" switch. cAMP can be visualized as the "on" switch which activates a number of kinases, and, as we will learn later,

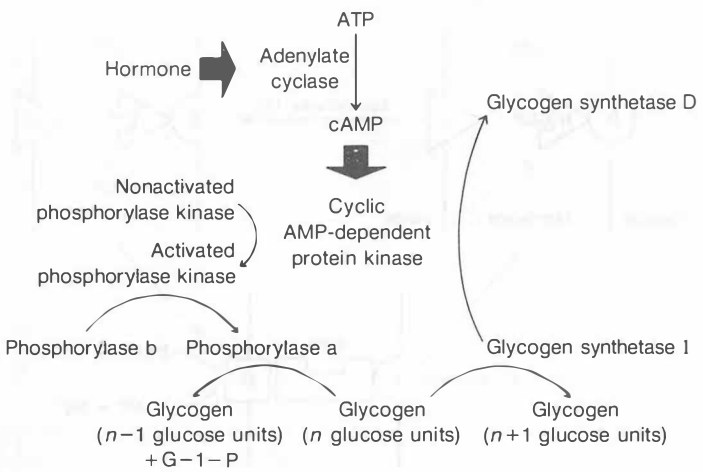


Figure 4.42
Hormonal activation of the membrane-bound adenylate cyclase produces the allosteric effector cAMP, which stimulates the protein kinases.
Protein kinase-catalyzed phosphorylation of glycogen synthetase I and phosphorylase kinase results in counterregulation of the synthesis and degradation of glycogen, that is, the phosphorylated synthetase D is less active, but the phosphorylated phosphorylase a is active. The series of enzymatic activations leading to phosphorylase a is called an enzymatic "cascade."
Reprinted, with permission, from T. Soderling, J. Hickenbottom, E. Reimann, et al., *J. Biol. Chem.*, 245:6327, 1970.

causes the de novo synthesis of some specific enzymes (page 758). cAMP levels are controlled in two ways. As the hormone diffuses away from the membrane receptor, guanosine triphosphate (GTP) on the regulatory subunit of the adenylate cyclase is hydrolyzed to GDP. The GDP binds tightly to the regulatory protein and induces a conformational change which results in dissociation of regulatory and catalytic subunits of adenylate cyclase. As a result of the dissociation of the catalytic and regulatory subunits the catalytic subunit assumes an inactive conformation. These effects are summarized in Figure 4.43. Binding of additional hormone stimulates the release of the GDP and binding of GTP, which activates the adenylate cyclase. Cytoplasmic levels of cAMP also are controlled by a cAMP-dependent phosphodiesterase that catalyzes the reaction

$$\text{cAMP} + \text{H}_2\text{O} \longrightarrow \text{AMP} + \text{P}_\text{i}$$

thus removing the activator of the protein kinase cascade. This phosphodiesterase is subject to control by caffeine and related purines. These substances inhibit the diesterase, resulting in a stimulatory effect by allowing cAMP levels to remain above steady state concentrations.

Calmodulin

It now appears that calcium is another "second messenger" that functions through activation of a regulatory subunit of a number of cAMP-independent protein kinases. Calmodulin is a term coined for the "calcium-dependent regulatory protein." In the presence of micromolar amounts of calcium, calmodulin undergoes a conforma-

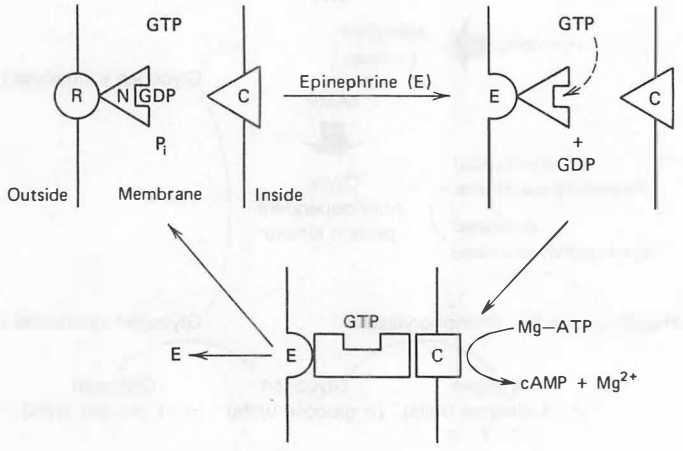


Figure 4.43
Model of hormone activation of adenylate cyclase.
The hormone receptor (R) is bound to the regulator (N) that has bound to it guanosine diphosphate (GDP). The GDP serves as a "stop" signal in the absence of hormone. Binding of a hormone like epinephrine triggers the release of GDP, which allows GTP to bind. GTP serves as an effector that initiates conformational changes in the regulator subunit so that adenylate cyclase catalytic subunit (C) is activated. Diffusion of epinephrine from the receptor allows the hydrolysis of GTP to GDP resulting in reversal of the conformation of adenylate cyclase to the inactive form.

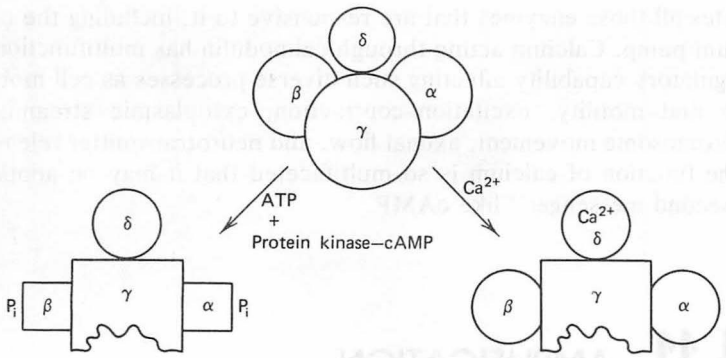


Figure 4.44
Modulation of the phosphorylase kinase reaction by cAMP-dependent protein kinase and calcium activated calmodulin.
The catalytic subunit is γ. The α and β subunits are phosphorylated by protein kinase, which results in activation. Another route of activation is through the δ subunit, which is calmodulin. Calcium binding to calmodulin activates the catalytic subunit of phosphorylase kinase.

tional change, which is transmitted cooperatively to the catalytic subunit of a number of specific enzyme kinases, such as phosphorylase kinase and glycogen synthase kinase. Calmodulin is a calcium-dependent protein allosteric activator. In phosphorylase kinase, the α and β subunits are phosphorylated by a cAMP-dependent protein kinase, while the δ subunit is identical to calmodulin. The γ subunit is the catalytic subunit whose activity is regulated both by cAMP and calcium levels as suggested in Figure 4.44. Although the scheme shown in Figure 4.44 illustrates the control of phosphorylase kinase in particular through both cAMP and calcium via calmodulin, the mechanism may be much more general because it allows for control by two independent mediators, which would result in fine tuning of the metabolic pathway.

The cAMP phosphodiesterase also contains calmodulin as the regulatory protein. Calcium increases phosphodiesterase activity; hence the intracellular cAMP level is controlled indirectly by the flux of calcium in the cell.

Calcium and calmodulin function is much broader than would be indicated by our discussion of the phosphorylase kinase and phosphodiesterase enzymes. In muscle the intracellular calcium level is controlled by an ATP dependent "pump," which responds indirectly to nervous stimulation. Stimulation increases the intracellular calcium levels. The calcium binds to calmodulin, which then acti-

vates all those enzymes that are responsive to it, including the calcium pump. Calcium acting through calmodulin has multifunctional regulatory capability affecting such diverse processes as cell mobility and motility, excitation–contraction, cytoplasmic streaming, chromosome movement, axonal flow, and neurotransmitter release. The function of calcium is so multifaceted that it may be another "second messenger" like cAMP.

4.11 AMPLIFICATION OF REGULATORY SIGNALS

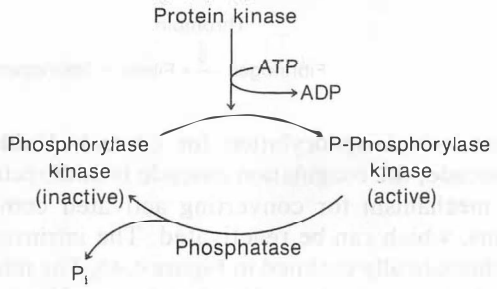
In biological systems many of the signal molecules, such as hormones, that function in interorgan communication are very low in concentration but have tremendous effects on the target organ. The signal generated by the binding of very small amounts of hormone is multiplied manifold inside the cell through a process of biological amplification. The mechanism of this amplification involves a cascade of reactions whereby the activation of the initial enzyme in the cascade activates a second proenzyme, and it in turn activates a third proenzyme, and so on. Since catalytic proteins are involved at each step in the cascade, the initial signal can be increased many times in terms of amounts of the final product generated.

Phosphorylase Cascade

A beautiful example of amplification is the epinephrine-stimulated phosphorolysis of glycogen as depicted in Figure 4.42 (left-hand reaction series). Levels of epinephrine on the order of 10^{-10} mol/g of muscle will stimulate the formation of 25×10^{-6} moles of glucose 1-phosphate per minute per gram of muscle. This is an amplification factor of 250,000. If we look at the initial stages of the amplification, we find that the epinephrine raises the steady-state cAMP concentration only three- to fivefold. This may at first appear to be a rather small amplification, but we must remember that the hormone is binding to only a limited number of receptor sites on the cell surface, which only affects the activation of an equivalent number of adenylate cyclase molecules. In addition, phosphodiesterases rapidly hydrolyze the cAMP formed, therefore we are not able to measure the full extent of amplification. Theoretical considerations have shown that the logarithm of the concentration of effector needed to activate 50% of the ultimate target enzymes is inversely proportional to the

number of steps in the cascade. Since there are three steps between cAMP and the activation of phosphorylase, it has been calculated that only a 1% increase in the steady-state level of cAMP would produce a 50% activation of phosphorylase; consequently, three- to fivefold increase in cAMP is theoretically sufficient to activate all the phosphorylase.

The phosphorylase cascade in Figure 4.42 is depicted as a unidirectional series of reactions; however, in practice each step is composed of a series of closed bicyclic phosphorylations and dephosphorylation reactions. For example



The advantage of such a bicyclic system is that the potential for control as well as amplification is tremendous; that is, each kinase and phosphatase could be subject to different metabolic allosteric modifiers at each step in the cascade.

In general those factors that limit the extent of amplification in a system involving a catalytic enzyme cascade are the relative amounts and the turnover numbers of each enzyme in the cascade. If each step in the cascade is bicyclic, then the ratio of the forward rate (activation) to the backward rate (deactivation) is important. If the activated enzyme is inactivated by inhibitors, as occurs in the blood coagulation cascade discussed below, then the relative rates of activation and inactivation become important.

It is interesting that the concentrations of the enzymes in the phosphorylase cascade are of the ratio expected for an amplification system. In fast twitch muscles, the molar ratios of cAMP-dependent protein kinase, phosphorylase kinase, and phosphorylase are 1 : 10 : 240. In slow twitch muscles the ratios are 1 : 0.15 : 25. Thus the absolute concentration as well as the ratio of the enzymes in a cascade are altered to provide the chemical response appropriate to the functioning of a particular tissue.

Blood Coagulation

Another cascade that is of great biological importance occurs in the coagulation of blood and involves a mechanism other than

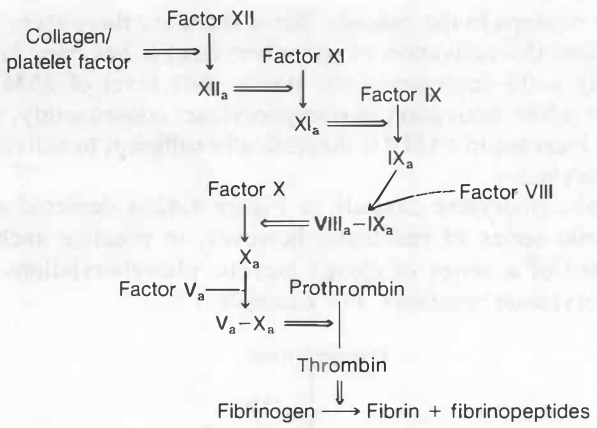


Figure 4.45
The intrinsic pathway of blood coagulation. *Each factor is a proenzyme and each activated factor is a specific peptidase. Unlike the phosphorylase cascade, activation of each subsequent enzyme in the cascade involves hydrolytic removal of peptide segments from the inactive factor. The end product of the cascade, thrombin, is inactivated by the plasma protein antithrombin III. Such plasma inhibitors provide a means of regulation of the coagulation cascade. The double arrows indicate the catalyzed phases of the reaction.*

phosphorylation–dephosphorylation for control. Unlike the phosphorylase cascade, the coagulation cascade is unidirectional, that is, there is no mechanism for converting activated components into inactive forms, which can be reactivated. The intrinsic coagulation scheme is schematically outlined in Figure 4.45. The intrinsic system contains all the components within the plasma. No tissue factor is required.

In this cascade the initiating agent converts Hageman Factor (Factor XII) to a specific endopeptidase, which then converts the proenzyme of Factor XI to the active endopeptidase, Factor XI_a, by hydrolytic removal of some peptides. Each subsequent activation also requires the conversion of a proenzyme to an active peptidase. Exposure of Factor XII to collagen or to stimulated platelets results in activation of Factor XII. As can be seen from Figure 4.45, there are six cycles of enzyme activation in the cascade. Amplification in this system must be on the order of 10^6. The terminal reaction in the pathway is the conversion of the soluble plasma protein fibrinogen to an insoluble polymer of fibrin. Thrombin catalyzes the hydrolysis of two acidic fibrinopeptides from the α and β chains of fibrinogen resulting in formation of fibrin monomers, which polymerize to form a fibrin clot.

The cascade is controlled by protein inhibitors of the activated enzymes of the pathway. These inhibitors are α_1-antitrypsin, antithrombin III, and α_2-macroglobulin. The most important is antithrombin III. Since there is an excess of inhibitor over activated thrombin, our blood does not clot completely. Clotting occurs in regions of locally high thrombin concentrations, such as at the site of a cut, because the rate of formation of the thrombin–inhibitor complex is slow compared to the rate of fibrin formation. The sulfated heteropolysaccharide, heparin, markedly enhances the rate of bind-

ing of antithrombin III to thrombin in the blood. Heparin is also a component of normal blood vessel wall, and presumably helps to prevent adventitious clot formation in that locale except following a cut or other trauma. Heparin may serve as an allosteric activator of the antithrombin protein. In addition to these effects of heparin, it also inhibits the activation of Factor IX by Factor XI_a and of Factor VIII by IX_a. Heparin therefore, indirectly controls the conversion of prothrombin to thrombin. Heparin is released into the blood by the mast cells of the blood vessels walls of the lung and liver, in particular. The levels of activated clotting factors are also controlled by the liver which removes the activated factors from the plasma and destroys them by enzymatic digestion.

Autoregulation of thrombin production is also a mechanism of control of the cascade. Thrombin has a high affinity for Factors VIII and V, and through its endopeptidase activity degrades these factors to forms that will not function in formation of the active IX_a–$VIII_a$ and X_a–V_a complexes. Thus the rate of thrombin formation becomes dependent upon the concentrations of Factor VIII and V, which in turn depend upon the concentration of thrombin.

BIBLIOGRAPHY

General Enzymology

Boyer, P. D., (ed.). *The enzymes*, 3rd ed. (13 vols.). New York: Academic, 1970–1976.

Nord, F. F. (ed.). *Advances in enzymology*. New York: Wiley-Interscience, issued annually.

Scrimgeour, K. G. *Chemistry and control of enzyme reactions*. New York: Academic, 1977.

Coenzymes

Chaykin, S. Nicotinamide coenzymes. *Annu. Rev. Biochem.* 36:149, 1967.

Gunsalus, I. C., Pederson, T. C., and Sligar, S. G. Oxygenase-catalyzed biological hydroxylations. *Annu. Rev. Biochem.* 44:377, 1975.

Krampitz, L. O. Catalytic functions of thiamin diphosphate. *Annu. Rev. Biochem.* 38:213, 1969.

Neims, A. H., and Hellerman, L. Flavoenzyme catalysis. *Annu. Rev. Biochem.* 39:867, 1970.

Kinetics

Cleland, W. W. Enzyme kinetics. *Annu. Rev. Biochem.* 36:77, 1967.

Piszkiwicz, D. *Kinetics of chemical and enzyme-catalyzed reactions.* New York: Oxford University Press, 1977.

Segel, I. H. *Enzyme kinetics*. New York: Wiley, 1975.

Nomenclature

Enzyme Nomenclature. Recommendations of the international union of biochemistry on the nomenclature and classification of enzymes, together with their units and symbols of enzyme kinetics. Elsevier, 1972.

Intracellular Distribution of Enzymes

DePierre, J. W., and Ernster, L. Enzyme topology of intracellular membranes. *Annu. Rev. Biochem.* 46:201, 1977.

Dingle, J. T. (ed.). *Lysosomes, a laboratory handbook.* Amsterdam: North-Holland, 1977.

Greer, D. E., and MacLenna, D. H. The mitochondrial system of enzymes, in D. M. Greenberg (ed.), *Metabolic pathways*, vol. I. New York: Academic, 1967.

Clinical Enzymology

Brewer, G. J., and Sing, C. F. *An introduction to isozyme techniques.* New York: Academic, 1970.

Coodley, E. L. *Diagnostic enzymology.* Philadelphia: Law & Febiger, 1970.

Davie, E. W., Fujikowa, K., Kurachi, K., et al. The role of serine proteases in the blood coagulation cascade, in D. Bodansky and A. L. Latner (eds.), *Adv. Clin. Chem.* 48:277, 1979.

Galen, R. S. Isoenzymes of CPK and LDH in myocardial infarction and certain other diseases. *Pathobiol. Annu.* 5:283, 1975.

Wilkinson, J. H. *Principles and practice of diagnostic enzymology.* Chicago: Year Book Publishers, 1976.

The Active Site and Mechanism of Enzyme Action

Crane, F. L. Hydroquinone dehydrogenases. *Annu. Rev. Biochem.* 46:439, 1977.

Koshland, D. E., Jr., and Neet, K. E. The catalytic and regulatory properties of enzymes. *Annu. Rev. Biochem.* 37:359, 1968.

Kraut, J. Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* 46:331, 1977.

Mildvan, A. S. Mechanisms of enzyme action. *Annu. Rev. Biochem.* 43:357, 1974.

Sigman, D. S., and Mooser, G. Chemical studies of enzyme active sites. *Annu. Rev. Biochem.* 44:889, 1975.

Walsh, C. Chemical approaches to the study of the enzymes catalyzing redox transformations. *Annu. Rev. Biochem.* 47:881, 1978.

Wimmer, M. J., and Rose, I. A. Mechanisms of enzyme-catalyzed group transfer reactions. *Annu. Rev. Biochem.* 47:1031, 1978.

Wood, H. G., and Barden, R. E. Biotin enzymes. *Annu. Rev. Biochem.* 46:385, 1977.

Regulation

Cohen, P. Hormonal control of glycogen metabolism in mammalian muscle by multivalent phosphorylation. *Biochem. Soc. Trans.* 7:459, 1979.

Gumaa, K. A., McLean, P., and Greenbaum, A. L. Compartmentation in relation to metabolic control in liver. *Essays Biochem.* 7:39, 1971.

Krebs, E. G., and Beavo, J. A. Phosphorylation-dephosphorylation of enzymes. *Annu. Rev. Biochem.* 48:923, 1979.

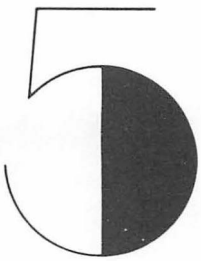
Kun, E., and Grisolia, S. *Biochemical regulatory mechanisms in eukaryotic cells.* New York: Wiley, 1972.

Schimke, R. T., and Doyle, D. Control of enzyme levels in animal tissues. *Annu. Rev. Biochem.* 39:929, 1970.

Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. (eds.). *The metabolic basis of inherited disease,* 4th ed. New York: McGraw-Hill, 1978.

Umbarger, H. E. Amino acid biosynthesis and its regulation. *Annu. Rev. Biochem.* 47:533, 1978.

Weber, G. (ed.). *Advances in enzyme regulation,* vols. 1–6. Elmsford, N.Y.: Pergamon, 1963–1979.



Biological Membranes

THOMAS M. DEVLIN

STRUCTURE AND MEMBRANE TRANSPORT

5.1 OVERVIEW

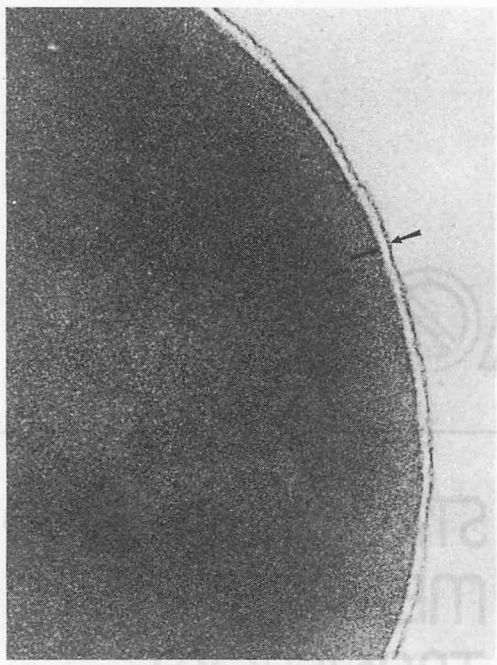


Figure 5.1
Electron micrograph of the erythrocyte plasma membrane showing the trilaminar appearance. *The two electron-dense lines separate a clear space. Electron microscopy has demonstrated that the inner dense line is frequently thicker than the outer line. Magnification about 150,000×; Courtesy of Dr. J. D. Robertson.*

As discussed in Chapter 1, a major difference between procaryotic and eucaryotic cells is the presence of a variety of membrane bound structures in eucaryotic cells. Regardless of the cell of origin, biological membranes, whether those that surround the cell or enclose an intracellular organelle, have a number of similar components, with a general structural organization common to all membranes. There are, of course, major differences in the specific lipid and protein components but not in the interaction of these components, which leads to the unique appearance and properties of membranes. All membranes can control the composition of the space they enclose not only by their ability to exclude a variety of molecules but also because of the presence of selective transport systems permitting the movement of specific molecules from one side to the other. By controlling the translocation of substrates, cofactors, ions, and so on, from one compartment to another, membranes modulate the concentration of substances, thereby exerting an influence on metabolic pathways. The plasma membrane of eucaryotic cells also has a role in cell–cell recognition, maintenance of the shape of the cell and in cell locomotion. The site of action of many hormones and metabolic regulators is on the plasma membrane (Chapter 16), where there are specific recognition sites, and the information to be imparted to the cell by the hormone or regulator is transmitted by the membrane to the appropriate metabolic pathway by a series of intracellular intermediates, termed second messengers.

All biological membranes have a trilaminar appearance when viewed by electron microscopy (Figure 5.1), with two dark bands on each side of a light band. The overall width of the various mammalian membranes is 7–10 nm; some membranes, however, have significantly smaller widths. Differences are also observed in both the size and the density of membranes depending on the staining technique employed in preparing tissue sections making it difficult to determine the significance of the differences in width. Intracellular membranes are usually thinner than the plasma membrane. In addition, many membranes do not appear symmetrical, with the inner dense layer often thicker than the outer dense layer; as discussed below there is a chemical asymmetry of the membrane, with some components only on one or the other side. With the development of sophisticated techniques for preparation of tissue samples and staining, including negative staining and freeze fracturing, the various surfaces of membranes have been viewed; at the molecular level the surfaces are not smooth but dotted with globular-shaped compo-

nents protruding from the membrane. Data from electron microscopic evaluation of membranes has been useful in developing current concepts on the molecular structure of membranes.

Even though valuable in defining structure, electron micrographs present a very static picture of membranes. Membranes are very dynamic structures with a movement that permits the cell as well as subcellular structures in eucaroytic cells to adjust their shape and to move. As reviewed below, the chemical components of membranes, that is, lipids and protein, are ideally suited for the dynamic role of membranes. Membranes are visualized as essentially a semi-structured, organized sea of lipid in a fluid state in which the various components are able to move. The fluid state of the lipid membrane is essentially a nonaqueous compartment of the cell where enzyme catalyzed reactions can occur in a controlled environment. Thus, in considering the role of membranes in the various activities of the cell, we should remember their dynamic state.

The discussion that follows is directed primarily to the chemistry and function of membranes of mammalian cells but the basic observations and functions described are applicable to all biological membranes.

5.2 CHEMICAL COMPOSITION OF MEMBRANES

Lipids and proteins are the two major components of all membranes but the percent of each varies greatly between different membranes (Figure 5.2). The percent of protein ranges from about 20% in the myelin sheath to over 70% in the inner membrane of the mitochondria. Intracellular membranes have a high percent of protein presumably because of the greater enzymatic activity of these membranes. Membranes also contain a small percent of various polysaccharides in the form of glycoprotein and glycolipid; there is no free carbohydrate in membranes.

Lipids of Membranes

The three major lipid components of membranes are phosphoglycerides, sphingolipids, and cholesterol; individual cellular membranes also contain small quantities of other lipids, such as triacylglycerol and diol derivatives, but their functions are uncertain. The percentage of each of the major classes varies significantly in different membranes and is presumably related to the specific

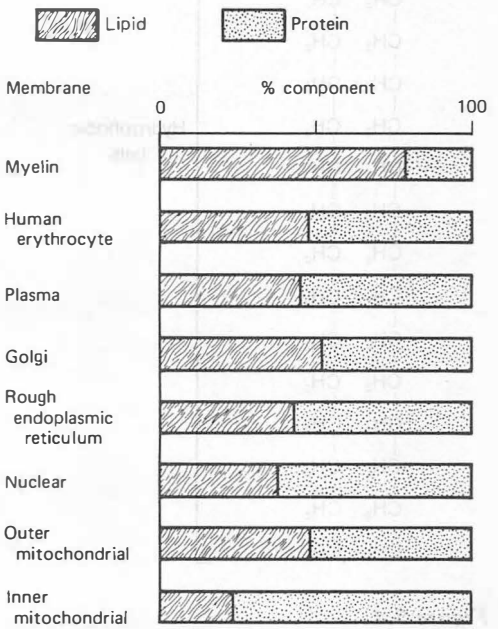


Figure 5.2
Representative values for the percent lipid and protein in various cellular membranes.
Values are for rat liver, except for the myelin and human erythrocyte plasma membrane. Values for liver from other species, including human, indicate a similar pattern.

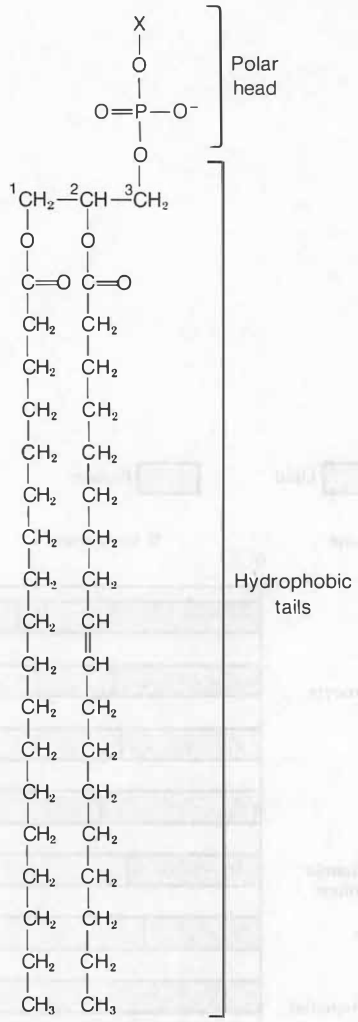


Figure 5.4
Structure of phosphoglyceride.
Long-chain fatty acids are esterified at C-1 and C-2 of the L*-glycerol 3-phosphate. X can be a H (phosphatidic acid) or one of several alcohols presented in Figure 5.5.*

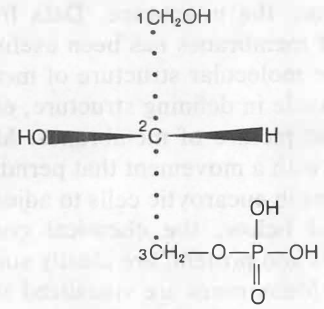


Figure 5.3
Stereochemical configuration of L***-glycerol 3-phosphate (sn-glycerol 3-phosphate).***
The H and OH attached to C-2 are above and C-1 and C-2 are below the plane of the page.

roles of the individual membranes. This is discussed in more detail below.

Phosphoglycerides

Phosphoglycerides, also referred to as glycerophospholipids, have a glycerol molecule as the basic component to which phosphoric acid is esterified at the α-carbon (Figure 5.3) and two long-chain fatty acids are esterified at the remaining carbons (Figure 5.4). Even though glycerol does not contain an asymmetric carbon, the α-carbons are not stereochemically identical. Esterification of a phosphate to a carbon makes the molecule asymmetric. The naturally occurring phosphoglycerides are designated by the stereospecific numbering system (sn) as presented in Figure 5.3 and also discussed in Chapter 10, Section 10.2.

Phosphatidic acid, 1,2-diacylglycerol 3-phosphate, is the parent compound of a series of phosphoglycerides, where different hydroxyl containing compounds are also esterified to the phosphate groups. The major compounds attached by the phosphodiester bridge to glycerol are choline, ethanolamine, serine, glycerol, and inositol. These structures are presented in Figure 5.5. Phosphatidylethanolamine (also called ethanolamine phosphoglyceride and the trivial name cephalin) and phosphatidylcholine (choline phosphoglyceride or lecithin) are the two most common phosphoglycerides in membranes (Figure 5.6). Phosphatidylglycerol phosphoglyceride (Figure 5.7) (or diphosphatidylglycerol or cardiolipin) contains two phosphatidic acids linked by a glycerol and is found nearly exclusively in the inner membrane of mitochondria and in bacterial membranes.

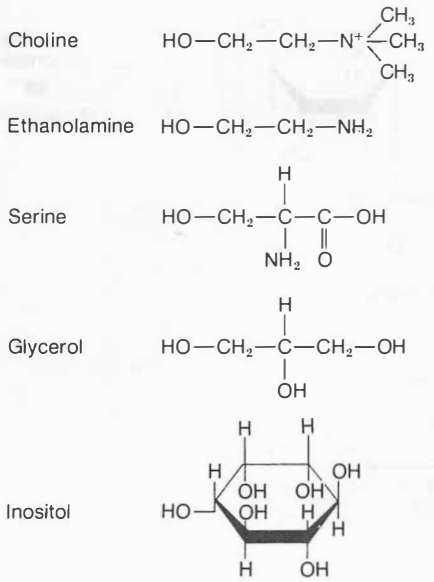


Figure 5.5
Structures of the major alcohols esterified to phosphatidic acid to form the phosphoglycerides.

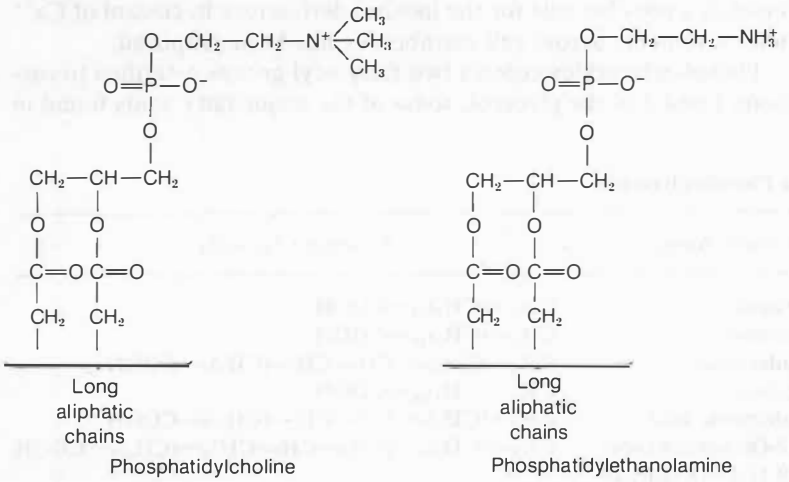


Figure 5.6
Structures of the two most common phosphoglycerides, phosphatidylcholine and phosphatidylethanolamine.

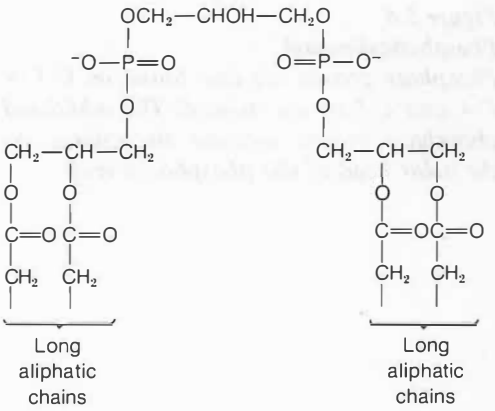


Figure 5.7
Diphosphatidylglycerol (cardiolipin).

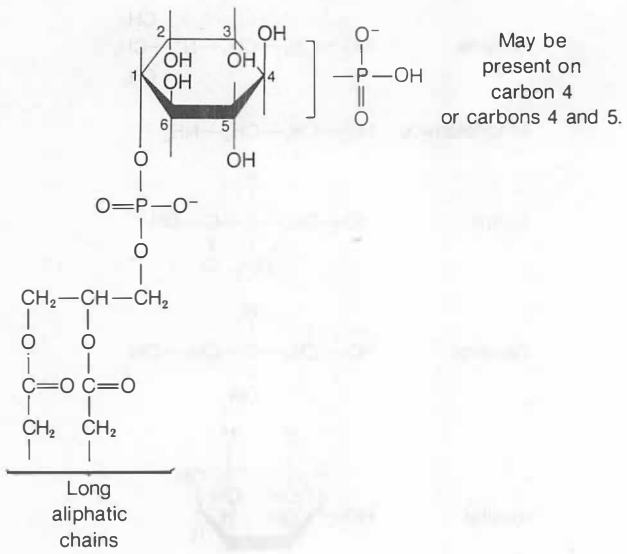


Figure 5.8
Phosphatidylinositol.
Phosphate groups are also found on C-4 or C-4 and C-5 of the inositol. The additional phosphate groups increase the charge on the polar head of this phosphoglyceride.

The hexahydroxy alcohol inositol is esterified to the phosphate in phosphatidylinositol (Figure 5.8); this compound should be differentiated from a class of lipids termed glycosylacylglycerols which contain a sugar in glycosidic linkage with the 3-hydroxyl group of a diacylglycerol. 4-Phospho- and 4,5-bisphosphoinositol phosphoglycerides (Figure 5.8) are found in various tissues, particularly myelin; a possible role for the inositol derivatives in control of Ca^{2+} ion movements across cell membranes has been proposed.

Phosphoglycerides contain two fatty acyl groups esterified to carbons 1 and 2 of the glycerol; some of the major fatty acids found in

Table 5.1 Major Fatty Acids in Phosphoglycerides

Common Name	*Systematic Name*	*Structural Formula*
Myristic acid	*n*-Tetradecanoic	$CH_3—(CH_2)_{12}—COOH$
Palmitic acid	*n*-Hexadecanoic	$CH_3—(CH_2)_{14}—COOH$
Palmitoleic acid	*cis*-9-Hexadecenoic	$CH_3—(CH_2)_5—CH=CH—(CH_2)_7—COOH$
Stearic acid	*n*-Octadecanoic	$CH_3—(CH_2)_{16}—COOH$
Oleic acid	*cis*-9-Octadecenoic acid	$CH_3—(CH_2)_7—CH=CH—(CH_2)_7—COOH$
Linoleic acid	*cis-cis*-9,12-Octadecadienoic	$CH_3—(CH_2)_3—(CH_2—CH=CH)_2—(CH_2)_7—COOH$
Linolenic acid	*cis,cis,cis*-9,12,15-Octadecatrienoic	$CH_3—(CH_2—CH=CH)_3—(CH_2)_7—COOH$
Arachidonic acid	*cis,cis,cis,cis*-5,8,11,14-Icosatetraenoic	$CH_3—(CH_2)_3—(CH_2—CH=CH)_4—(CH_2)_3—COOH$

phosphoglycerides are presented in Table 5.1. A saturated fatty acid is often found on carbon-1 of the glycerol and an unsaturated fatty acid on carbon-2. The nomenclature for the subclasses of phosphoglycerides does not specify a specific compound because of the variety of possible fatty acid substitutions. Phosphatidylcholine usually contains palmitic or stearic in the sn-1 position and an 18-carbon unsaturated fatty acid, oleic, linoleic or linolenic, on the sn-2-carbon. Phosphatidylethanolamine also contains palmitic or oleic on sn-1 but one of the longer chain polyunsaturated fatty acids, that is, arachidonic, on the sn-2 position.

A saturated fatty acid is a straight chain, as is a fatty acid with an unsaturation in the *trans* position. The presence of a *cis* double bond, however, creates a kink in the hydrocarbon chain, which alters the degree of hydrophobic interaction of the acyl chains in a membrane (Figure 5.9).

Another group of phosphoglycerides are the plasmologens in which a long aliphatic chain is attached in ether linkage to the

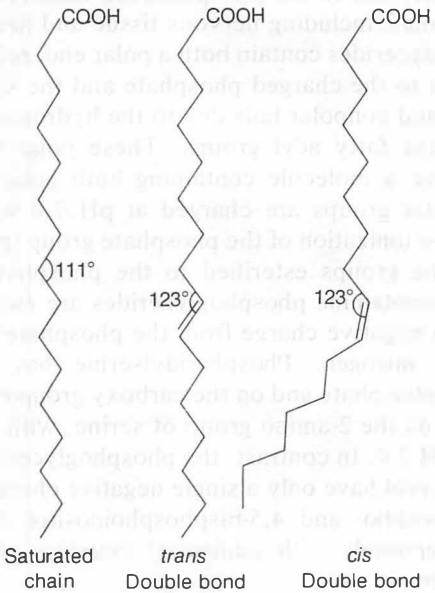


Figure 5.9
Conformation of fatty acyl groups in phospholipids.
The saturated and unsaturated fatty acids with trans *double bonds are straight chains in their minimum energy conformation, whereas a chain with a* cis *double bond has a bend. The* trans *double bond is rare in naturally occurring fatty acids.*

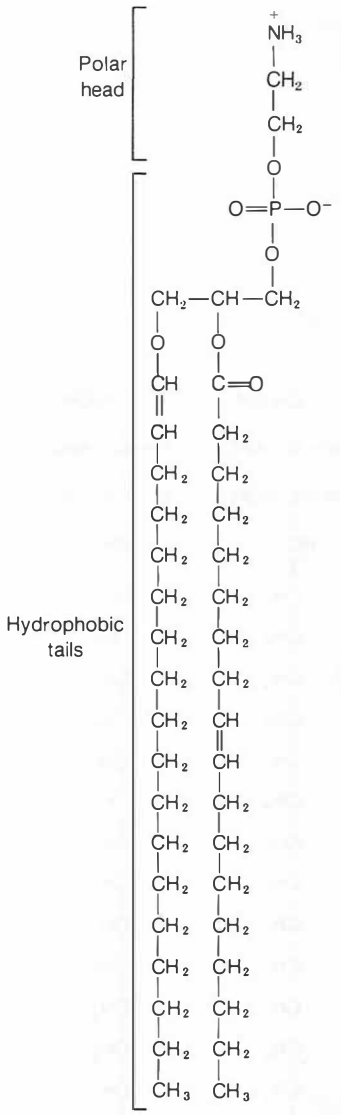


Figure 5.10
Ethanolamine plasmologen.
Note the ether linkage of the aliphatic chain on C-1 of glycerol.

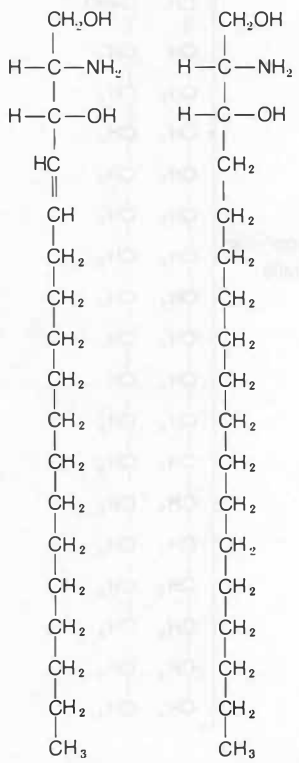


Figure 5.11
Structures of sphingosine and dihydrosphingosine.

Table 5.2 Predominant Charge on Phosphoglycerides and Sphingomyelin at pH 7.0

Lipid	*Phosphate Group*	*Base*	*Net Charge*
Phosphatidylcholine	−1	+1	0
Phosphatidylethanolamine	−1	+1	0
Phosphatidylserine	−1	+1, −1	−1
Phosphatidylglycerol	−1	0	−1
Diphosphatidylglycerol (cardiolipin)	−2	0	−2
Phosphatidylinositol	−1	0	−1
Sphingomyelin	−1	+1	0

glycerol as presented in Figure 5.10. Plasmalogens containing ethanolamine (ethanolamine plasmalogen) and choline (choline plasmalogen) esterified to the phosphate are relatively abundant in various membranes, including nervous tissue and heart.

The phosphoglycerides contain both a polar end, referred to as the head group, due to the charged phosphate and the substitutions on the phosphate, and nonpolar tails due to the hydrophobic hydrocarbon chains of the fatty acyl groups. These polar lipids are amphipathic, that is, a molecule containing both polar and nonpolar groups. The polar groups are charged at pH 7.0 with a negative charge due to the ionization of the phosphate group ($pK \sim 2$) and the charges from the groups esterified to the phosphate (Table 5.2). Choline and ethanolamine phosphoglycerides are zwitterions at pH 7.0, with both a negative charge from the phosphate and a positive charge on the nitrogen. Phosphatidylserine has two negative charges, on the phosphate and on the carboxy group of serine, and a positive charge on the 2-amino group of serine, with a net negative charge of 1 at pH 7.0. In contrast, the phosphoglycerides containing inositol and glycerol have only a single negative charge on the phosphate; the 4-phospho- and 4,5-bisphosphoinositol derivatives are very polar compounds with additional negative charges on the phosphate groups.

Every tissue and respective cellular membrane has a distinctive composition of phosphoglycerides. Not only are there differences in the classes of phosphoglycerides, but there are definite patterns in the fatty acid composition of the individual phosphoglycerides between tissues. In fact, there appears to be some degree of specificity for particular fatty acids in the individual tissues. There is a greater variability in the fatty acid composition of different tissues in a single

species than the fatty acid composition of the same tissue in a variety of species.

Sphingolipids

The amino alcohols sphingosine (4-sphingenine) and dehydrosphingenine (Figure 5.11) serve as the basis for another series of membrane lipids, the sphingolipids. To the amino group of the sphingosine base a saturated or unsaturated long chain fatty acid is present in amide linkage. This compound, termed a ceramide, (Figure 5.12) with two nonpolar tails is similar in structure to the diacylglycerol portion of phosphoglycerides. Various substitutions

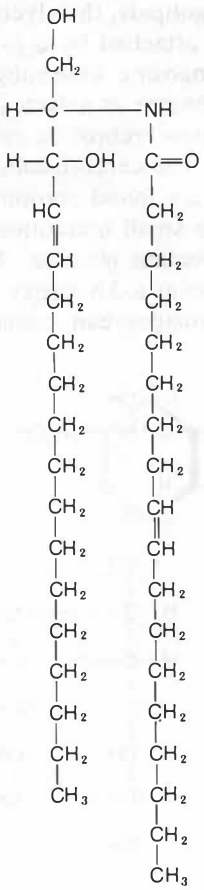


Figure 5.12
Structure of a ceramide.

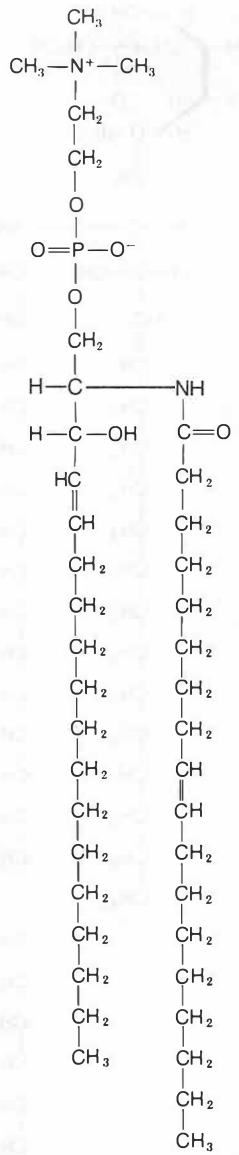


Figure 5.13
Structure of a choline containing sphingomyelin.

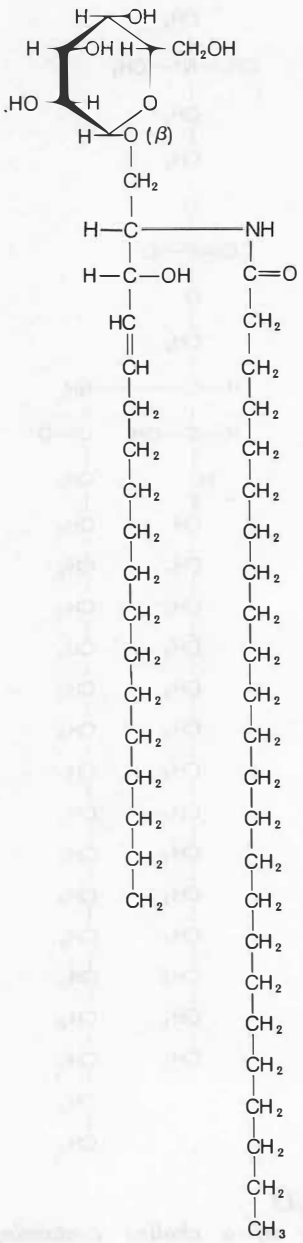


Figure 5.14
Structure of a galactocerebroside containing a C_{24} fatty acid.

are found on the hydroxyl group at position 1 of the ceramides. The sphingomyelin series has phosphorylcholine or phosphorylethanolamine esterified to the 1-hydroxyl (Figure 5.13) and are the most abundant sphingolipids in mammalian tissues. The similarity of these structures to the choline and ethanolamine phosphoglycerides is apparent, and they have many properties in common; note that the sphingomyelins are amphipathic compounds. It has been a common practice to classify the sphingomyelin series and the phosphoglycerides in one class of compounds, termed phospholipids. The sphingomyelin of myelin contains predominantly the longer chain fatty acids, with carbon lengths of 24; as with phosphoglycerides, there is a specific fatty acid composition of the sphingomyelin, depending on the tissue.

Another class of sphingolipids, the glycosphingolipids, has one or several sugar molecules attached by a β-glycosidic linkage to the 1-hydroxyl group of sphingosine. One subgroup is the cerebrosides, which contain either a glucose or galactose attached to a ceramide and are referred to as glucocerebroside and galactocerebroside, respectively (Figure 5.14). The cerebrosides are neutral compounds. The galactocerebrosides are found predominantly in brain and nervous tissue, whereas the small quantities of cerebrosides in nonneural tissues usually contain glucose. The specific galactocerebroside, phrenosine, contains a 2-hydroxy 24-carbon fatty acid (Figure 5.15). Galactocerebrosides can contain a sulfate group (sul-

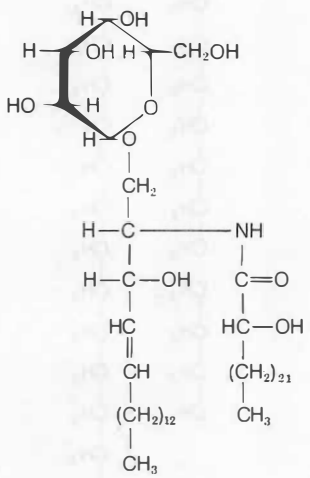


Figure 5.15
Structure of phrenosine.
Note the 2-hydroxy on the fatty acid.

fatides), as shown in Figure 5.16, esterified on the 3 position of the sugar, which is found in brain tissue. Cerebrosides and sulfatides usually contain very long-chain fatty acids with 22 to 26 carbon atoms.

In addition to containing monosaccharides, neutral glycosphingolipids often have 2, 3, and 4 sugars, termed dihexasides, trihexasides, and tetrahexasides, respectively, with diglucose, digalactose, *n*-acetylglucosamine, and *n*-acetyldigalactosamine as sugars.

The most complex group of glycosphingolipids are the gangliosides, which contain oligosaccharide head groups with one or more residues of sialic acid; these are amphipathic compounds with a negative charge at pH 7.0. The gangliosides represent 5–8% of the total lipids in brain, and some 20 different types have been identified differing in the number and relative position of the hexose and sialic acid residues, which form the basis of their classification. A detailed description of these structures is presented in Chapter 10.

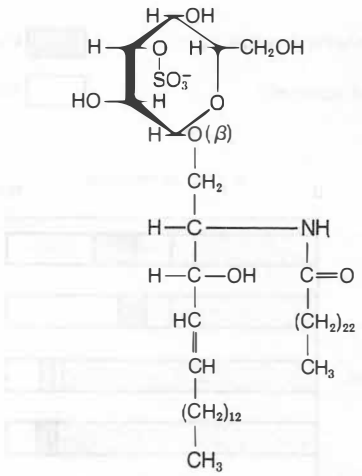


Figure 5.16
Structure of a sulfatide.

Cholesterol

The third major lipid present in membranes is cholesterol. As presented in Figure 5.17 cholesterol contains four fused rings, which makes it a planar structure, a polar hydroxyl group at carbon-3, and an eight-member branched hydrocarbon chain attached to the D ring at position 17. Cholesterol is a compact hydrophobic molecule.

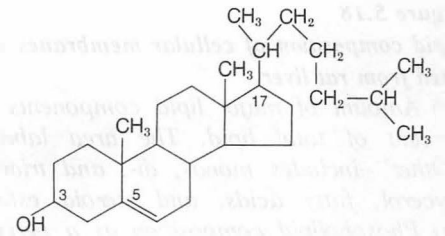


Figure 5.17
Structure of cholesterol.

Distribution of Membrane Lipids

There are large quantitative differences in both classes of lipids and individual lipids in various cell membranes. Figure 5.18 presents the lipid composition of various cellular membranes.

There is a resemblance between species in the lipid composition of the same intracellular membrane of cells in specific tissue, that is, mitochondria of liver of rat and man. The plasma membrane exhibits the greatest variation in percent composition because the amount of cholesterol is affected by the nutritional state of the animal. Plasma membranes have the highest concentration of neutral and sphingolipids; the myelin membranes of axons of neural tissue are rich in sphingolipids, with a high proportion of glycosphingolipids. Intracellular membranes contain primarily phosphoglycerides with little or no sphingolipids or cholesterol. When comparing intracellular structures, the membrane lipid composition of mitochondria, nuclei, and rough endoplasmic reticulum are similar, with the Golgi membrane being somewhere between the other intracellular membranes and the plasma membrane. As indicated previously, cardiolipin is found nearly exclusively in the inner mitochondrial membrane. The

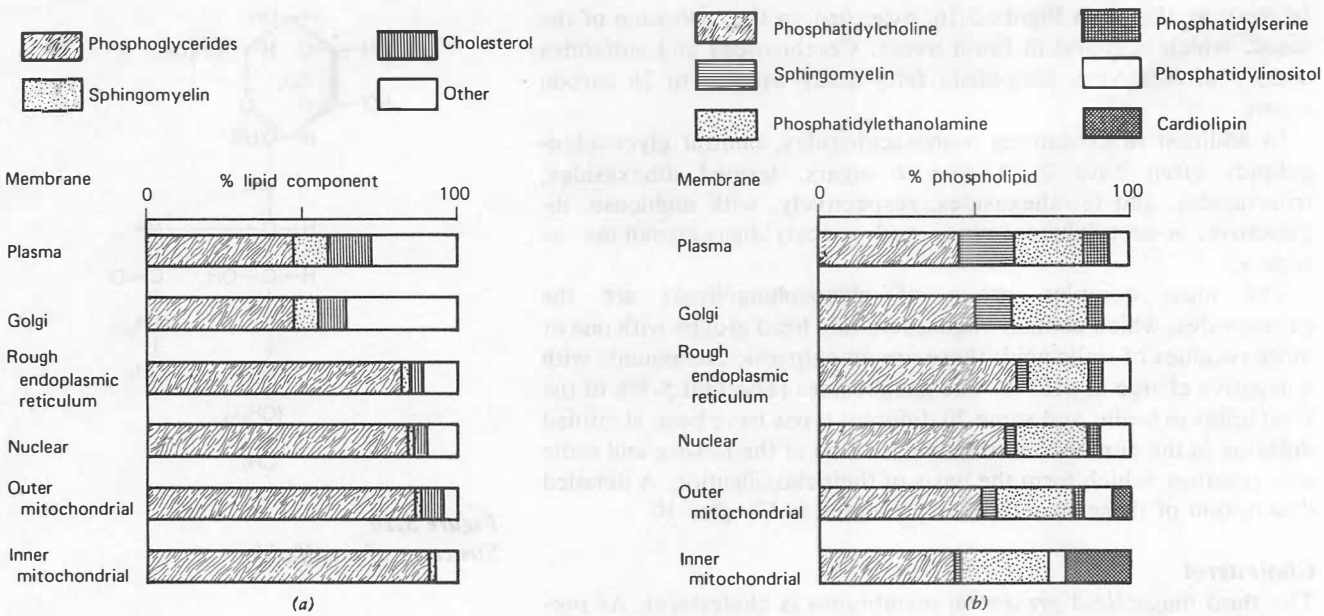


Figure 5.18
Lipid composition of cellular membranes isolated from rat liver.
(*a*) *Amount of major lipid components as percent of total lipid. The area labeled "Other" includes mono-, di-, and triacylglycerol, fatty acids, and steroid esters.* (*b*) *Phospholipid composition as a percent of total phospholipid.*

Values from R. Harrison and G. G. Lunt, *Biological Membranes*, Wiley, New York, 1975.

choline containing lipids, phosphatidylcholine and choline ceramide are predominant, with ethanolamine phosphoglyceride second. The constancy of composition of the various membranes indicates the relationship between the lipids and the specific functions of the individual membranes.

Proteins of Membranes

Membrane proteins are classified on the basis of the ease of removal from isolated membrane fractions. Extrinsic (or peripheral) proteins are easily isolated by treatment of the membrane with salt solutions of low or high ionic strength, or extremes of pH, and the name is used to imply a physical location on the surface of the membrane. Extrinsic proteins, many with specific enzymatic activity, are usually soluble in water and free of lipids. Intrinsic (or integral) proteins require rather drastic treatment, such as use of detergents, to be extracted from the membrane. Intrinsic proteins are usually lipoprotein in nature and are not soluble in water. Removal of the intrinsic protein leads to disruption of the membrane, whereas extrinsic proteins can be removed with little or no change in the integrity of the membrane.

Of particular value in studying the chemistry and structure of intrinsic proteins has been the use of sodium dodecyl sufate (SDS), a

detergent which dissociates the lipid protein complex and solubilizes the protein permitting separation and analysis. The intrinsic proteins studied have sequences of hydrophobic amino acids, which could create domains with a high degree of hydrophobicity in the tertiary structure of the protein. These hydrophobic regions of the protein interact with the hydrophobic hydrocarbons of the lipids stabilizing the protein–lipid complex.

A special class of intrinsic proteins are the proteolipids, which are hydrophobic lipoproteins soluble in chloroform and methanol but insoluble in water. Proteolipids are present in many membranes but are particularly abundant in myelin, where they represent about 50% of the membrane protein component.

Plasma membranes contain a significant quantity of carbohydrate in the form of glycoprotein. It has been found that most of the carbohydrate is on the exterior side of the plasma membrane and is essentially outside the cell. One role for the membrane glycoprotein is in cell recognition. Glycoproteins do occur in nearly all membranes but usually in amounts much smaller than the plasma membrane.

The complexity, variety, and interaction of membrane proteins with lipids are just being resolved. Many of the proteins are enzymes located within or on the cellular membranes. Membrane proteins also have a role in transmembrane movement of molecules and in many cells, such as neurons and erythrocytes, specific proteins have a structural role to maintain the integrity of the cell. Thus individual membrane proteins can have a catalytic, transport, structural, or recognition role, and it is not surprising to find a high protein content in a membrane being correlated with complexity and extent of function of the membrane.

5.3 MOLECULAR STRUCTURE OF MEMBRANES

Micelles, Membranes, and Liposomes

The basic structural characteristic of all membranes is derived from the physicochemical properties of the major lipid components, the phosphoglycerides and sphingolipids. These amphipathic compounds, with a hydrophilic head and a hydrophobic tail, react in a unique fashion in an aqueous system because of their very low solubility in water. Above a specified concentration, referred to as the critical micelle concentration, the hydrophobic tails interact to ex-

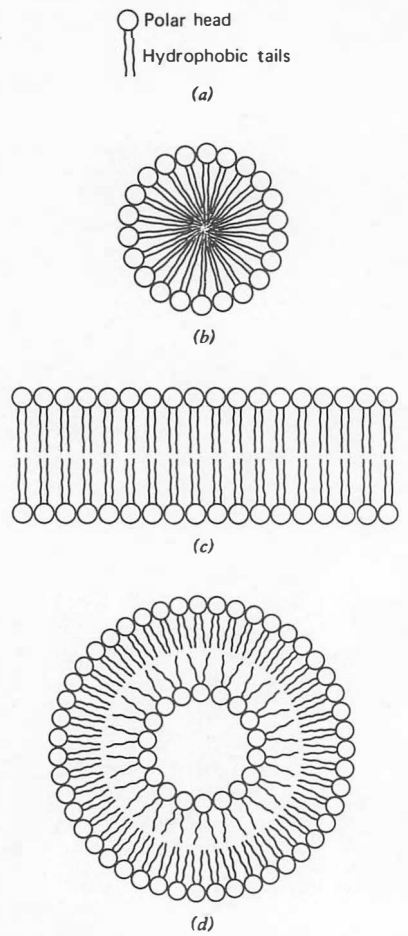


Figure 5.19
Representations of the interactions of phospholipids in an aqueous medium.
(a) Representation of an amphipathic lipid. (b) Cross-sectional view of the structure of a micelle. (c) Cross-sectional view of the structure of lipid bilayer. (d) Cross section of a liposome. Each structure has an inherent stability due to the hydrophobic interaction of the hydrocarbon chains and the attraction of the polar head groups to water.

clude water and form spheres, termed micelles, with the polar head groups on the outside as visualized in Figure 5.19. Micelles with various lipid composition can be formed. The formation of the micelle depends also on the temperature of the system and, if a mixture of lipids are used, on the ratio of concentrations of the different lipids in the mixture. The micelle structure is very stable because of the hydrophobic interaction of the hydrocarbon chains and the attraction of the polar groups to water. As discussed in Chapter 24, Section 24.6, micelles are important in the digestion of lipids.

Under special conditions, amphipathic lipids interact to form a bimolecular leaf structure with two layers of lipid in which the polar head groups are at the interface between the aqueous medium and the lipid and the hydrophobic tails interact to form an environment that excludes water (Figure 5.19). This bilayer conformation is the basic lipid structure of all biological membranes.

Lipid bilayers are extremely stable structures held together by a number of noncovalent interactions. Hydrophobic interaction of the hydrocarbon chains leads to the smallest possible area for water to be in contact with the hydrophobic groups; water is essentially excluded from the interior of the bilayer. The interaction between water and the polar groups leads to structures that resist disruption, and the bilayers are self-sealing because the hydrophobic portion will seek the structure where there is the least contact of water with the acyl chains, a condition that is most thermodynamically favorable. A lipid bilayer will close in on itself, forming a vesicle separating the external space from an internal compartment. These vesicles are termed liposomes. Because the individual lipid–lipid interactions have low energies of activation, the lipids in a bilayer have a circumscribed mobility, breaking and forming interactions with surrounding molecules but not readily escaping from the lipid bilayer. The bilayer has not only an inherent stability but also a fluidity in which individual molecules can move in the structure. In artificial bilayer membranes composed of different lipids, the components will be randomly distributed.

Artificial membrane systems have been studied extensively as a means to determine the properties of biological membranes. A variety of techniques are available to prepare liposomes, using synthetic phospholipids and lipids extracted from natural membranes. Depending on the procedure, unilamellar vesicles and multilamellar vesicles (vesicles within vesicles) of various sizes can be prepared. Figure 5.19 contains a representation of the structure of a liposome. The interior of the vesicle is an aqueous environment, and it is possible to prepare liposomes with different substances entrapped. Thus the external and internal environment of the liposome can be

manipulated and studies conducted on a variety of properties of these synthetic membranes, including their ability to exclude molecules, their interaction with various substances, and their stability under different conditions. Na^+, K^+, Cl^-, and most polar molecules do not readily diffuse across the lipid bilayer of liposomes, whereas the membrane presents no barrier to water. Lipid-soluble nonpolar substances such as triacylglyceride and undissociated organic acids readily diffuse into the membrane remaining in the hydrophobic environment of the hydrocarbon chains. Proteins, both synthetic and natural, have been incorporated into the membrane to mimic the natural membrane. Membrane-bound enzymes and proteins involved in translocating ions have been isolated from various tissues and incorporated into the membrane of liposomes for evaluation of the proteins function. With the liposome it is easier to manipulate the various parameters of the membrane system and, thus, study the catalytic activity free of possible interfering reactions that are present in the natural membrane. Liposomes have also been used in drug therapy (Clin. Corr. 5.1).

Biological Membranes

Based on evidence from physicochemical, biochemical, and electron microscopic investigations, knowledge of the structure of biological membranes has evolved. The basic structure is a bimolecular leaf arrangement of lipids in which the phosphoglycerides, sphingolipids, and cholesterol are oriented so that the hydrophobic portions of the molecules interact to minimize their interactions with water or other polar groups. The polar head groups of the amphipathic compounds are at the interface with the aqueous environment. This arrangement of lipid is similar to that in synthetic phospholipid liposomes. A major area of investigation has been directed to explaining the interaction of the intrinsic and extrinsic membrane proteins in the lipid bilayer. A number of models for the structure of biological membranes have been proposed dating back to one proposed by H. Davson and J. Danielle in 1935, which was refined in later years by J. D. Robertson. In the 1960s, G. L. Nicolson and S. J. Singer proposed the *mosaic* model for membranes, in which the proteins are on the surface of, embedded in, and even transverse the lipid bilayer, with the polar groups on both proteins and lipid in contact with the aqueous surroundings and hydrophobic portions of both in the interior of the membrane. This model has been extensively refined and, to indicate the apparent movement of both lipids and proteins in the membrane, is referred to as the *fluid mosaic* model. Figure 5.20 is a pictorial representation of a biological membrane as proposed by Singer and Nicolson. The proposed structure accounts for many of

CLIN. CORR. 5.1 LIPOSOMES AS CARRIERS OF DRUGS AND ENZYMES

A major obstacle in the use of many drugs is the lack of tissue specificity in the action of the drug. Drugs administered orally or intravenously often lead to the drug acting on many tissues and not exclusively on a target organ and is the basis for their toxicity. In addition, some drugs are metabolized rapidly and their period of effectiveness is relatively short. Liposomes have been prepared with drugs and even enzymes encapsulated inside and used as carriers for these substances to target organs. Liposomes prepared from purified phospholipids and cholesterol are nontoxic and biodegradable. Attempts have been made to prepare liposomes for interaction at a specific target organ. Some drugs have a longer period of effectiveness when administered encapsulated in liposomes. It is hoped, as our understanding of the structure and chemistry of cellular membranes increases, that it will be possible to prepare liposomes with a high degree of tissue specificity so that drugs and perhaps even enzyme replacement can be carried out with this technique. (For a brief review of the topic, see G. Gregoriadis, *New Engl. J. Med.* 295:704, 1976.)

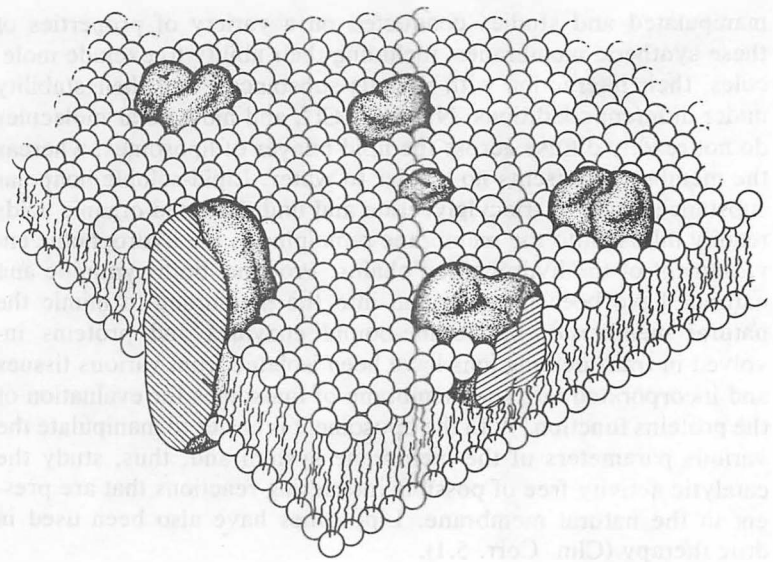


Figure 5.20
The fluid mosaic model of biological membranes.
The membrane consists of a fluid phospholipid bilayer with globular integral proteins penetrating the bilayer.

Reproduced with permission from S. J. Singer and G. L. Nicolson, *Science*, 175:720, 1972. Copyrighted 1972 by the American Association for the Advancement of Science.

the properties of mammalian membranes, but it needs to be emphasized that the model is under continuous modification. There are many observations which are not readily explained by the model presented in Figure 5.20. The physicochemical properties of the lipid bilayer, however, do account for many membrane properties, including the flexibility of cellular membranes, their self-sealing properties, and the impermeability of many substances.

An important difference between the Nicolson–Singer model and earlier models is the proposal that the lipid bilayer is discontinuous with proteins actually embedded in the hydrophobic portion of the bilayer. This has been confirmed, and in fact many membrane proteins span the bilayer with portions protruding on each side. Hydrophobic interaction between the lipid and hydrophobic domains on these intrinsic proteins prevents these proteins from being readily removed, and their extraction leads to disruption of the membrane. Obviously, there are significant differences in the strengths of interaction of different intrinsic proteins and the lipid bilayer.

Proteins are also bound to the membrane by electrostatic interaction between charges on the protein and the charged groups of the amphipathic lipids. The proteins on the surface of the membrane are extrinsic proteins, removed easily by mild treatment with little damage to the membrane. Even though the model would suggest that proteins are randomly distributed throughout the membrane, evidence from a variety of sources suggests a high degree of functional organization with definite restrictions on the localization. As an ex-

ample, proteins participating in electron transport in the inner-membrane of mitochondria function in consort and are organized into a functional unit both laterally and traversely in the membrane. Membrane proteins also have a definite orientation within and across the membrane; the forces that direct this detailed organization are not known. Many intrinsic proteins with enzymatic activity have their catalytic site on either the inner or the outer surface, and peripheral proteins are bound specifically to only one side. As an example, the carbohydrate moieties of plasma membrane glycoproteins are found on the outer surface. Thus there is a high degree of molecular organization of biological membranes which is not necessarily apparent from the diagrammatic models.

Not only is there an asymmetry across the membrane with respect to proteins, but the lipid components are also distributed asymmetrically. Each layer of the bilayer has a different composition with respect to individual phosphoglycerides and sphingolipids. The asymmetric distribution of lipids in erthryocyte membranes is presented in Figure 5.21. Note that sphingomyelin is predominantly in the outer layer of lipids, whereas phosphatidylethanolamine is predominantly in the inner lipid layer. Similar results have been observed for a number of membranes. How the cell directs and maintains the asymmetric distribution of lipids in membranes is unknown, but it is believed that this distribution is important in membrane function.

The interactions between the different lipids and between lipids and proteins are very complex and dynamic. There is a fluidity to the lipid portion of the membrane in which both the lipids and proteins move. The degree of fluidity is dependent on temperature; at lower temperatures the lipids are in a gel crystalline state, and as the temperature is increased a transition into a liquid crystalline state with an increase in fluidity occurs. The transition temperature for biological membranes is not precise, in that it depends on the composition of the individual membranes and the interactions between the hydrocarbon chains, cholesterol, other lipid molecules, and proteins in the membrane. In contrast, the transition temperature for synthetic liposomes prepared from pure phospholipids can be determined relatively precisely.

The fluidity at different levels within the membrane will also vary. The hydrocarbon chains have a molecular motion, which produces a fluidity in the hydrophobic core. The very central area occupied by the ends of the hydrocarbon chains is presumably more fluid than the areas closer to the two surfaces, where there are more constraints due to the stiffer portions of the hydrocarbon chains. Cholesterol with its flat stiff ring structure makes the membrane more rigid toward the periphery because it does not reach into the central core of

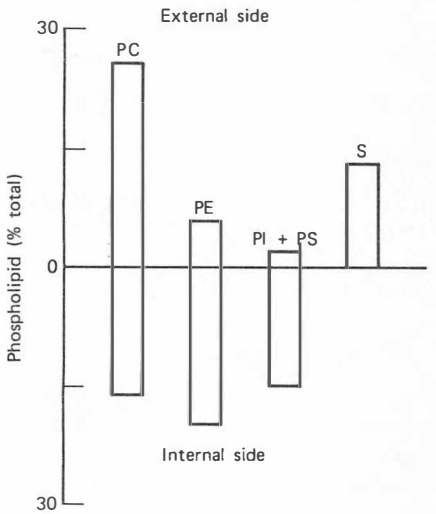


Figure 5.21
Proposed distribution of phospholipids between inner and outer layers of the rat erythrocyte membrane.
Values are percent of each phospholipid in the membrane. Abbreviations: PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; PS, phosphatidylserine; S, sphingomyelin.
Data from W. Renooij, L. M. G. van Golde, R. F. A. Zwaal, and L. L. M. van Deenen, *Eur. J. Biochem.*, 61:53, 1976.

the membrane. The presence of the *cis* conformation of double bonds in the fatty acids prevents tight packing of the chains and creates pockets in the hydrophobic areas. It is assumed that these spaces, which will also be mobile due to the mobility of the hydrocarbon chains, are filled with water molecules and small ions.

Individual lipids and proteins move in a lateral motion along the surface of the membrane. The lateral movement of individual lipids is very rapid and is measured in minutes, whereas transverse movement from one side to the other (i.e., flipflop movement) is very slow and is measured in days or weeks. The slow rate of transverse movement is not unexpected, considering how unfavorable in thermodynamic terms it is to force the hydrophilic polar head group of a phospholipid through the hydrophobic interior of a membrane and then reorient the group on the opposite side. The asymmetry of lipids in the erythrocyte membrane is an example of how slow is the transverse movement of membrane lipids. The mature erythrocyte has a lifetime of about 120 days, during which time there is no new membrane synthesis or even significant repair. Even so, there appears to be little mixing of the phospholipids between the molecular layers. There are restrictions also on the lateral movement of lipids; electrostatic interactions of polar headgroups, hydrophobic interactions of cholesterol with selected phospholipids or glycolipids, and protein–lipid interactions all lead to constraints on the movement.

Intrinsic and extrinsic membrane proteins also move in the lipid environment. A demonstration of the lateral movement has come from fusing human and rat cells after antigenic membrane proteins on cells of each species were prelabeled with a different antibody marker. The marker permitted the localization of the two different proteins on the membrane. Immediately following fusion of the cells, proteins of the human and rat were segregated in different hemispheres of the new cell, but within 40 min the two proteins were evenly distributed over the new cell membrane. This could only occur if there had been lateral diffusion of the proteins on the cell membrane. Protein motion is slower than that of the lipids. There is either no or very little transverse motion of proteins in membranes. Some membrane proteins immobilize a ring of lipid around the protein, which restricts the movement of both the protein and the surrounding lipid molecules. Intrinsic membrane proteins may also be restricted by other membrane proteins, matrix proteins, or cellular structural elements such as filaments and microtubules.

Another aspect of the dynamic state of membrane lipids and proteins is their turnover. Individual lipids exchange with lipids in the cell matrix, as well as with lipids of other membranes. Specific mechanisms to maintain both the composition and asymmetry of the membrane must exist. Individual membrane proteins are being con-

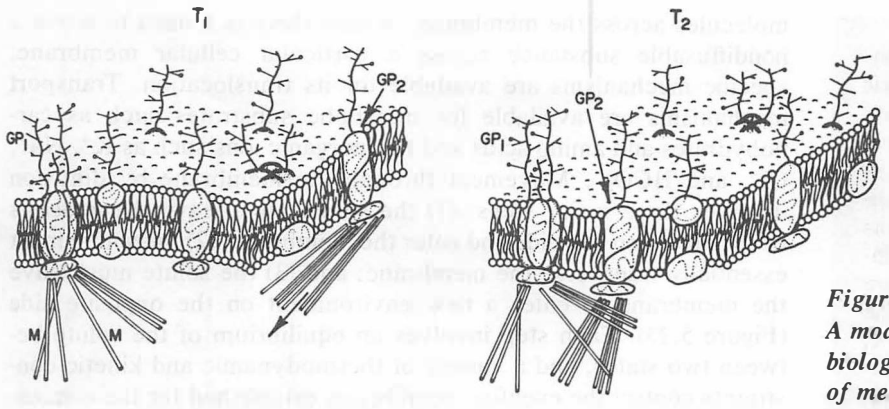


Figure 5.22
A modified version of the fluid mosaic model of biological membranes to indicate the mobility of membrane proteins.
T_1 and T_2 represent different points in time. Some integral proteins (GP_2) are free to diffuse laterally in the plane of the membrane directed by the cytoskeletal components, whereas others (GP_1) may be restricted in their mobility.
Reproduced with permission from J. L. Nicolson, *Int. Rev. Cytol.*, 39:89, 1974.

stantly removed and replaced by newly synthesized proteins, participating in the dynamic synthesis and degradation, which occurs with nearly all cellular components.

Thus cellular membranes are in a constantly changing state, with not only movement of proteins and lipids laterally on the membrane but with molecules moving into and out of the membrane. A variety of forces, including hydrophobic and electrostatic interactions, is involved in maintaining the basic structural characteristics. The membrane creates a number of microenvironments, from the hydrophobic portion of the core of the membrane to the interface with the surrounding environments. It is difficult to express in words or pictures the very fluid and dynamic state, in that neither captures the time-dependent changes that occur in the structure of biological membranes. Figure 5.22 attempts to illustrate the structural and movement aspects of cellular membranes (see Clin. Corr. 5.2).

5.4 MOVEMENT OF MOLECULES THROUGH MEMBRANES

The lipid nature of biological membranes severely restricts the type of molecules that will readily diffuse from one side to another. Substrates carrying a charge, whether inorganic ions or charged organic molecules, will not diffuse at a significant rate because of the attraction of these molecules to water molecules and the exclusion of the charged species by the hydrophobic environment of the lipid membrane. The diffusion rate of such molecules, however, is not zero but may be too slow to accommodate a cellular need to move these

CLIN. CORR. 5.2 ABNORMALITIES OF CELL MEMBRANE FLUIDITY IN DISEASE STATES

Changes in membrane fluidity can control the activity of membrane-bound enzymes, membrane functions such as phagocytosis, and cell growth. A major factor in controlling plasma membrane fluidity is the concentration of cholesterol. Higher organisms and mammals have a significant concentration of cholesterol in their membranes, which presumably has a major role in controlling the fluidity of the lipid

bilayer. With increasing cholesterol content membranes become less fluid on their outer surface but more fluid in the hydrophobic core. Individuals with spur cell anemia have an increased cholesterol content of the red cell membrane. This condition occurs in severe liver disease such as cirrhosis of the liver in alcoholics. Erythrocytes have a spiculated shape and are destroyed prematurely in the spleen. The cholesterol content is increased 25–65%, and the fluidity of the membrane is decreased. The erythrocyte membrane requires a high degree of fluidity for its function and any decrease would have serious effects on the cell's physiological role of oxygen transport. The increased plasma membrane cholesterol in other cells leads to an increase in intracellular membrane cholesterol, which also affects the fluidity of other cellular membranes. Individuals with abetalipoproteinanemia have an increase in sphingomyelin content and a decrease in phosphotidylcholine, thus causing a decrease in fluidity. The ramifications of these changes in fluidity are still not understood, but it is presumed that, as techniques for the measurement and evaluation of cellular membrane fluidity improve, some of the pathological manifestations in disease states will be explained on the basis of changes in membrane structure and function. (For further discussion see R. A. Cooper, Abnormalities of cell membrane fluidity in the pathogenesis of disease, *New Engl. J. Med.* 297:371, 1977.)

molecules across the membrane. Where there is a need to move a nondiffusable substance across a particular cellular membrane, specific mechanisms are available for its translocation. Transport mechanisms are available for metabolic substrates such as carbohydrates and amino acids and for inorganic ions such as K^+, Na^+, Cl^-, and HCO_3^-. Movement through the membrane by diffusion involves three major steps: (1) the solute must leave the aqueous environment on one side and enter the membrane; (2) the solute must essentially transverse the membrane; and (3) the solute must leave the membrane to enter a new environment on the opposite side (Figure 5.23). Each step involves an equilibrium of the solute between two states, and a variety of thermodynamic and kinetic constraints control the eventual equilibrium established for the concentrations of the substance on the two sides of the membrane and the rate at which it can attain the equilibrium. For simple diffusion of a hydrophilic solute with strong interaction with water molecules, the solute must have the shell of water stripped away to enter the lipid milieu but regains it on leaving the membrane. For hydrophobic substances, the distribution between the aqueous phase and lipid membrane will depend on the degree of lipid solubility of the substance; very lipid-soluble materials will essentially dissolve in the membrane.

In mammalian systems, the transport mechanism for various substances involve intrinsic membrane proteins, which interact with the molecule to be transported very much like an enzyme reacts with its substrate. Evaluation of the rate of transport is considered in the same terms, as described in Chapter 4, concerning enzyme catalyzed reactions, and the substances translocated are referred to as substrates. Some cellular transport systems move substances against a concentration gradient, that is, move from a lower to a higher concentration, which is possible only if there is an expenditure by the cell of some form of energy.

The following discussion describes the mechanisms by which molecules cross various cellular membranes; examples of specific systems will be described for illustrative purposes, but throughout this book individual systems are described in the context of specific metabolic processes.

Diffusion Across Cellular Membranes

The rate of diffusion of a solute is directly proportional to its lipid solubility and diffusion coefficient in lipids; the latter is a function of the size and shape of the substance. Uncharged lipophilic molecules, for example, fatty acids and steroids, diffuse relatively rapidly but water-soluble substances, for example, sugars and inorganic ions,

diffuse very slowly. Water diffuses readily through biological membranes. At one time it was believed that membranes had pores, that is, protein lined channels through the membrane, whose size excluded most molecules except water; but definitive evidence for such pores is lacking. Movement of water is believed to occur via the gaps in the hydrophobic environment created by the random movement of the fatty acyl chains of the lipids. Water and other small polar molecules can move into these transitory spaces and equilibrate across the membrane from one gap to another.

The direction of movement of solutes in diffusion is always from a higher to a lower concentration and the rate is described by Fick's first law of diffusion:

$$J = -D\left(\frac{\delta c}{\delta x}\right)$$

where J = net amount of substance moved per time, D = the diffusion coefficient, and $\delta c/\delta x$ = the chemical gradient of the substance. As the concentration of solute on one side of the membrane is increased, there will be an increasing initial rate of diffusion as illustrated in Figure 5.24. A net movement of molecules from one compartment to another will continue until the concentration in each is at a chemical equilibrium. After equilibrium is attained, there will be a continued exchange of solute molecules from one side to another, but no net accumulation on one side can occur because this would recreate a concentration gradient.

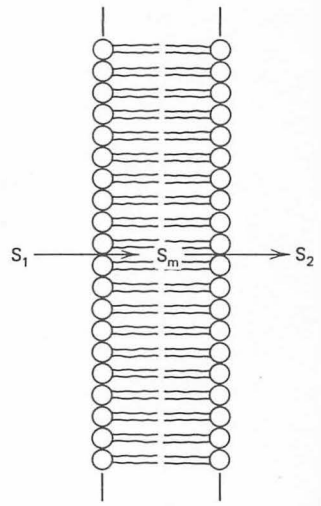


Figure 5.23
Diffusion of a solute molecule through a membrane.

Mediated Transport

Movement of a number of different substances through different cell membranes is facilitated by the presence of specific transport systems. Mechanisms are available for the movement of inorganic anions and cations (e.g., Na^+, K^+, Ca^{2+}, HPO_4^{2-}, Cl^-, and HCO_3^-) and uncharged and charged organic compounds (e.g., amino acids and sugars). Note that all cellular membranes do not have the same capability to move all substances; as an example, the plasma membrane has a mechanism to move K^+ and Na^+, which is not present in other cell membranes. The transport systems of mammalian cells involve integral membrane proteins with a high degree of specificity for the substances transported. These proteins or protein complexes have been designated by a variety of names including transporter, translocase, translocator, porter, and permease or termed transporter system, translocation mechanism, and mediated transport system to name a few. For some, the term pump is applied, but this is not a very descriptive term for membrane transport systems. The

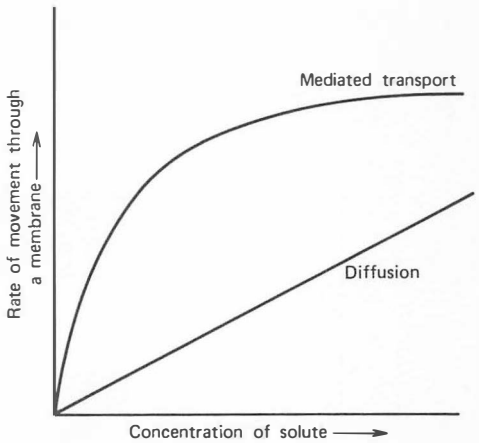


Figure 5.24
Kinetics of movement of a solute molecule through a membrane.
The initial rate of diffusion is directly proportional to the concentration of the solute. In mediated transport, the rate will reach a V_{max} *when the carrier is saturated.*

designations above are used interchangeably, but for convenience we will use transporter or translocase in referring to the actual protein involved in the translocation of a substance. In many cases the mediated movement of a substance can be measured, but the actual protein involved has not been studied; in these cases it is safer to refer to them as systems or mechanisms.

Membrane transporters have a number of characteristics in common. Each facilitates the movement of a molecule or molecules through the lipid bilayer at a rate which is significantly faster than can be accounted for by simple diffusion. If S_1 is the solute on side 1 and S_2 on side 2, then the transporter promotes an equilibrium to be established as follows:

$$[S_1] \rightleftharpoons [S_2]$$

where the brackets represent the concentration of solute. If the transporter (T) is included in the equilibrium the reaction is

$$[S_1] + T \rightleftharpoons [S_1 - T] \rightleftharpoons [S_2] + T$$

If there is no energy input by the system, the concentration on both sides of the membrane will be equal at equilibrium, but if there is an expenditure of energy, a concentration gradient can be established, which will depend on the thermodynamic properties of the system. Note the similarity of the role of the transporter to that of an enzyme, which increases the rate of a chemical reaction but does not determine the final equilibrium.

Mediated transport systems like enzyme-catalyzed reactions demonstrate saturation kinetics; as the concentration of the substance to be translocated increases, the initial rate of transport increases but reaches a maximum when the substance saturates the protein transporter on the membrane. A plot of solute concentration against initial rate of transport is hyperbolic, as presented in Figure 5.24. Simple diffusion does not demonstrate saturation kinetics. Constants such as V_{max} and K_m can be calculated for transporters as is done in studying an enzyme. Also, as with enzymes, transporters catalyze movement of a solute in both directions across the membrane.

Most transporters have a high degree of structural and stereospecificity for the substance transported. An example is the mediated transport system for D-glucose in the erythrocyte, where the K_m is 10 times larger for D-galactose than for D-glucose and for L-glucose it is 1,000 times larger. The transporter has essentially no activity with D-fructose or disaccharides. For many of the translocase systems competitive and noncompetitive inhibitors have been found. Struc-

tural analogs of the substrate inhibit competitively and reagents that react with specific groups on proteins are noncompetitive inhibitors.

The properties of saturation kinetics, substrate specificity, and inhibitability of transporters are characteristics in common with enzymes, but this information does not explain how the transporter actually facilitates the movement of a molecule across a distance in space. In considering mediated transport systems, therefore, we need to expand the equation above and consider four aspects: (1) recognition by the transporter of the appropriate solute from a variety of solutes in the aqueous environment, (2) translocation of the solute across the membrane, (3) release of the solute by the transporter, and (4) recovery of the transporter to its original condition to accept another solute molecule. This can be represented as follows, where T_1 is the transporter on side 1 and T_2 the transporter on side 2:

$$\begin{aligned}
&\text{Recognition:} && S_1 + T_1 \rightleftharpoons S\text{–}T_1 \\
&\text{Transport:} && S\text{–}T_1 \rightleftharpoons S\text{–}T_2 \\
&\text{Release:} && S\text{–}T_2 \rightleftharpoons T_2 + S_2 \\
&\text{Recovery:} && T_2 \rightleftharpoons T_1
\end{aligned}$$

The first step, recognition, can be readily explained on the same basis as that described for recognition of a substrate by an enzyme. The presence of very specific binding sites on the protein permits the transporter to recognize the correct structure of the solute to be translocated. The second step, translocation, is not understood. Models have been proposed, based on studies on different transporters for the actual movement of the solute, but none has received universal support. Diagrams of mediated transport systems frequently imply that a protein transporter, after binding the substrate, diffuses from one side of the membrane to the other, where the substrate is released (Figure 5.25*a*). This would require the tertiary structure of the protein to have a hydrophobic exterior in order to diffuse through the lipid milieu. Simple diffusion of a transporter substrate complex is unlikely, in that intrinsic membrane proteins do not move in a transverse direction in membranes, and it would be

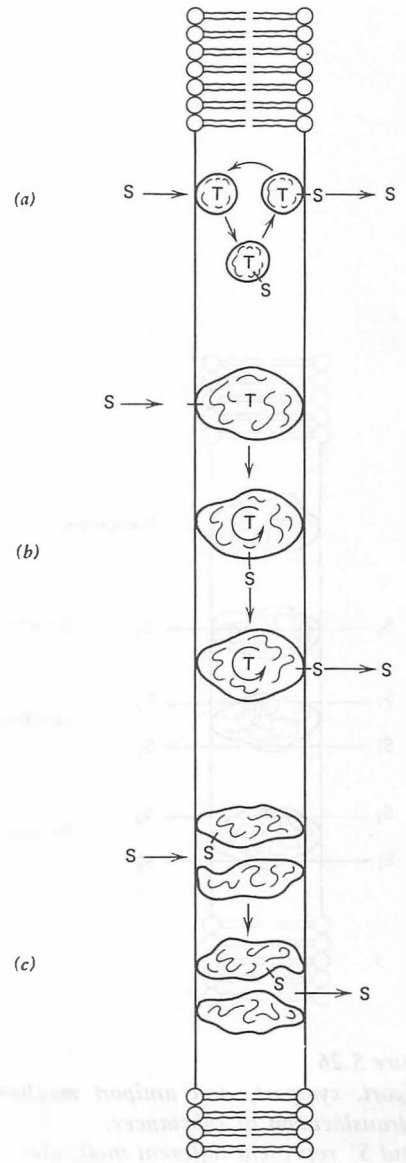


Figure 5.25
Models for mediated transport systems in biological membranes.
(a) The diffusion model where transporter diffuses back and forth across the lipid bilayer. (b) The rotation model where transporter rotates in the membrane. (c) The gated channel model where conformational changes in the transporter move the solute only a short distance but into the environment of the other side of the membrane.

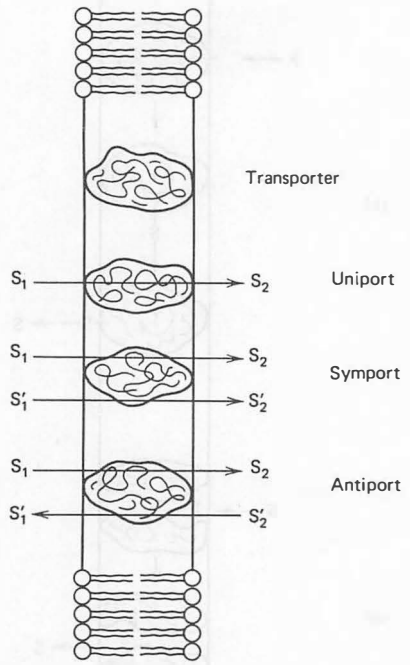


Figure 5.26
Uniport, symport, and antiport mechanisms for translocation of substances.
S and S' represent different molecules.

expected that the rate of diffusion of a large protein molecule would be very slow. Such a mechanism, however, is feasible, with relatively low molecular weight lipid-soluble compounds such as ionophores (Section 5.7) or several of the prostaglandins (Chapter 10, Section 10.5), which facilitate the movement of inorganic cations across membranes. Another possibility is for the translocase to rotate in the membrane (Figure 5.25*b*) accepting and releasing the solute on opposite sides of the membrane. This too is unlikely in that it would require the rotation of a rather bulky protein.

A reasonable model is one in which the protein transporter creates a link or channel between the environments on each side of the membrane with access through the channel being controlled by a gating mechanism (porter) in order to control which solutes can move into the channel. The transporter could have receptor sites to which the solute attaches. After association of the solute and translocase a conformational change of the protein could move the solute molecule a short distance, perhaps only 2 or 3 Å, but into the new environment of the opposite side of the membrane. In this manner, it is not necessary for the transporter physically to move the molecule the whole distance across the membrane (Figure 5.25*c*).

Release of the solute can occur readily based on simple equilibrium considerations if the concentration of solute is lower in the new compartment than on the initial side of binding. This does not require the dissociation constant of transporter–solute complex to change. For those translocases that move a solute against a concentration gradient it is possible to envision release of the solute at the higher concentration if the affinity for the solute by the transporter is decreased. A change in the conformation of the transporter could decrease the affinity. An alternate mechanism is to alter the solute chemically while attached to the translocase, so that it is a different molecule with a lower affinity for the transporter. Examples of both of these are described below.

Finally, in some way the transporter must return to its original state. If a conformational change has occurred, there must be a return to the original conformation.

The discussion above has centered on the movement of a single solute molecule by the transporter. Actually systems are known that move two molecules simultaneously in one direction (symport mechanisms), two molecules in opposite directions (antiport mechanism), as well as a single molecule in one direction (uniport mechanism) (Figure 5.26). Some transporters move charged molecules such as K^+, Na^+, and organic ions in which there is no direct and simultaneous movement of an ion of the opposite charge. The transporter creates a charge separation across the membrane, and the mechanism is termed electrogenic. If a counterion is moved to bal-

ance the charge, the mechanism is called neutral or electrically silent.

Energetics of Transport Systems

The change in free energy when an uncharged molecule moves from a concentration of C_1 to a concentration of C_2 on the other side of a membrane is given by the equation

$$\Delta G' = 2.3RT \log \frac{C_2}{C_1}$$

When $\Delta G'$ is negative, that is, release of free energy, the movement of solute will occur without the need for a driving force. When $\Delta G'$ is positive, as would be the case if C_2 is larger than C_1, then there needs to be an input of energy to drive the transport. For a charged molecule (e.g., Na^+) both the electrical potential and concentrations of solute are involved in calculating the change in free energy as follows:

$$\Delta G' = 2.3RT \log \frac{C_2}{C_1} + ZF\Psi$$

where Z is the charge of the species moving, F is the Faraday (23.062 kcal V^{-1} mol^{-1}), and Ψ is the difference in electrical potential in volts across the membrane. The electrical component is the membrane potential and $\Delta G'$ is the electrochemical potential.

A passive transport system is one in which $\Delta G'$ is negative, that is, free energy is released, and the movement of solute occurs spontaneously. When $\Delta G'$ is positive, coupled input of energy from some source is required for movement of the solute and the process is called active transport. Several different forms of energy are available for driving active transport systems, including hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP), and the electrochemical gradient of the Na^+ ions or of H^+ ions across the membrane. In the first the chemical energy released on hydrolysis of a pyrophosphate bond drives the reaction, whereas in the latter the electrochemical gradient is dissipated to transport the solute.

Transport systems that can maintain very large concentration gradients are present in various membranes. An example is the plasma membrane transport system which maintains the Na^+ and K^+ gradients. One of the most striking examples of an active transport system is that present in the parietal cells of gastric glands which are responsible for secretion of HCl into the lumen of the stomach (see Chapter 24, Section 24.3). The pH of plasma is about 7.4 (4×10^{-8}

M H^+), and the luminal pH of the stomach can reach 0.8 (0.15 M H^+). The cells transport H^+ ion against a concentration gradient of $1 \times 10^{6.6}$. Assuming there is no electrical component, the energy for H^+ ion secretion under these conditions can be calculated from the equation:

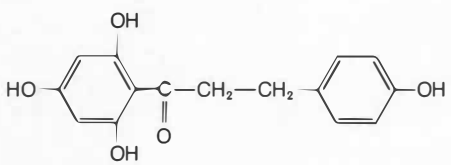
$$\Delta G' = 2.3RT \log 10^{6.6}$$

The energy required is 9.1 kcal/mol of HCl.

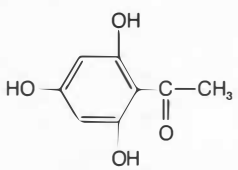
5.5 PASSIVE-MEDIATED TRANSPORT SYSTEMS

Passive-mediated transport, also referred to as facilitated diffusion, translocates solutes through cell membranes without the expenditure of metabolic energy. As with nonmediated diffusion the direction of flow is always from a higher to a lower concentration. The distinguishing differences between measurements of simple diffusion and passive-mediated transport are the demonstration of saturation kinetics, a structural specificity for the class of molecule moving across the membrane and specific inhibition of solute movement.

The plasma membrane of many mammalian cells, but not all, have a passive-mediated transport system for D-glucose. Most of our knowledge about this system is derived from studies of erythrocytes, particularly from humans. The physiological direction of movement is into the cell because the extracellular level of glucose is about 5 mM and most cells metabolize glucose rapidly thus maintaining low intracellular concentrations. Transport is by a uniport mechanism, which demonstrates saturation kinetics and is inhibitable. The system is most active with D-glucose, but D-galactose, D-mannose, and D-arabinose, and several other D sugars as well as glycerol can be translocated by the same transporter. The L-isomers are not transported. It has been proposed that the β-D-glucopyranose is transported with carrier interaction at the hydrogens on at least C_1, C_3, and C_6 of the sugar. The affinity of the erythrocyte carrier for D-glucose is highest with a K_m of about 6.2 mM, whereas for the other sugars the values are much higher. The carrier has a very low affinity for D-fructose, which precludes the carrier from having any role in cellular uptake of fructose. A separate carrier for fructose has been proposed. With isolated erythrocytes, glucose will move either into or out of the erythrocyte, depending on the direction of the experi-



Phloretin



2,4,6-Trihydroxyacetophenone

Figure 5.27
Inhibitors of passive mediated transport of D-glucose in erythrocytes.

mentally established concentration gradient, which demonstrates the reversibility of the system.

Several sugar analogues as well as phoretin and 2,4,6-trihydroxyacetophenone (Figure 5.27) are competitive inhibitors; reagents that react irreversibly with proteins are noncompetitive inhibitors. Treatment of cells from some tissues with the hormone insulin increases the V_{max} of glucose uptake; the effect of insulin is apparently indirect rather than directly on the carrier. The alteration in the carrier by the hormone demonstrates the probable importance of the carrier in maintenance of adequate intracellular levels of glucose.

Another passive-mediated transport system in erythrocytes involves the movement of the anions Cl^- and HCO_3^- by an antiport mechanism (Figure 5.28). In this mechanism Cl^- moves in one direction and simultaneously a HCO_3^- in the opposite direction; the direction of movement depends on the concentration gradients of the ions across the membrane. The transporter has an important role in adjusting the erythrocyte HCO_3^- concentrations in arterial and venous blood (see Chapter 23, Section 23.7) and is a part of the mechanism for removing CO_2 from the body.

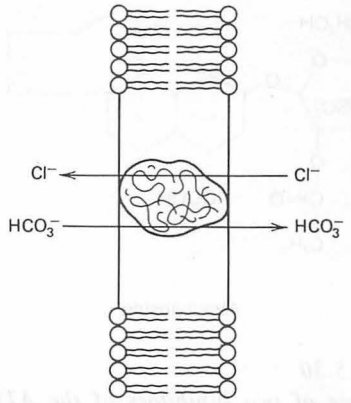


Figure 5.28
The passive antiport mechanism for movement of Cl^- and HCO_3^- across the erythrocyte plasma membrane.

The inner mitochondrial membrane contains several antiport systems for the exchange of anions between the cytosol and mitochondrial matrix. These include (1) a transporter for exchange of ADP and ATP; (2) for exchange of phosphate and OH^-; (3) a dicarboxylate carrier that catalyzes an exchange of malate for phosphate, and (4) a translocator for exchange of aspartate and glutamate (Figure 5.29). The relationship of these translocases and energy coupling are discussed in Chapter 6, Section 6.5. In the absence of an input of energy these transporters will catalyze a passive exchange of metabolites down their concentration gradient to achieve a thermodynamic equilibrium of all intermediates. As an antiport mechanism, a concentration gradient of one compound can drive the movement of the other solute. The systems, however, can also couple to the mitochondrial energy transducing system, and the anions are moved against their concentration gradient. The ability of these translocases to function as either a passive-mediated transport system or as an active-mediated transport system is unique.

The ATP–ADP translocase has been extensively studied; it is very

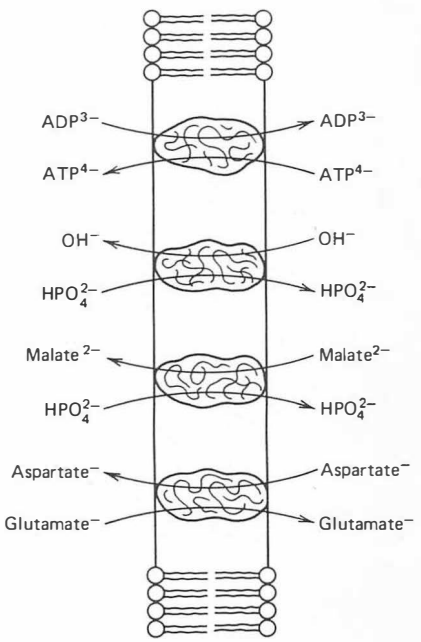
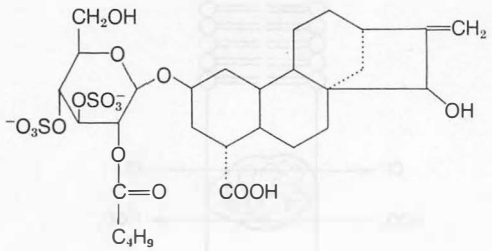
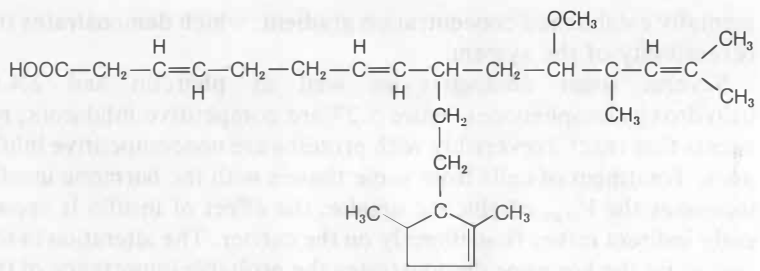


Figure 5.29
Representative anion transport systems in liver mitochondria.
Note that each is an antiport mechanism. Several other transport systems are known and are discussed in Chapter 6.



Atractyloside



Bongkrekic acid

Figure 5.30
Structure of two inhibitors of the ATP/ADP transport system of liver mitochondria.

specific for ATP and ADP and their deoxyribose derivatives, dATP and dADP, but does not transport adenosine monophosphate (AMP) and other nucleotides. The protein responsible for the translocation has been isolated and is a dimer containing two subunits of 30,000 daltons each. It has been estimated that the transporter represents 12% of the total protein in heart mitochondria. The protein is very hydrophobic and can exist in two conformations. Atractyloside and bongkrekic acid (Figure 5.30) are specific inhibitors, each apparently reacting with a different conformation of the protein. The mitochondrial membrane potential can drive the movement of the nucleotides by this translocator, but in the absence of the potential it will function as a passive-mediated transporter.

It is sometimes difficult to differentiate passive-mediated transport from simple diffusion, but specific inhibition is a good demonstration of a carrier; this has been the case for the anion carriers of mitochondria, which have been differentiated on the basis of specific inhibitors.

5.6 ACTIVE-MEDIATED TRANSPORT SYSTEMS

A variety of mediated transport systems requiring the utilization of energy to move solutes against their concentration gradient are present in the plasma membrane. Active transport systems, sometimes referred to as pumps, have the same three characteristics as passive transporters, that is, saturation kinetics, substrate specificity, and inhibitability, but in addition these systems require a coupled input of energy. If the energy source is removed or inhibited, the transport system will not function. Inhibition of ATP synthesis leads to the

inhibition of many transport systems, suggesting ATP as the primary energy source for active transport. Direct utilization of ATP, however, does not occur in all active transport systems. Several systems utilize the electrochemical gradient of Na^+ across the membrane. As indicated below in the discussion on the active-mediated transport of glucose, which utilizes the transmembrane Na^+ gradient in a symport mechanism, metabolic energy in the form of ATP is required for maintenance of the Na^+ gradient but not directly in moving a glucose molecule; inhibition of ATP synthesis, however, leads to a dissipation of the Na^+ electrochemical gradient, which in turn decreases transport activity utilizing the gradient. This is visualized in Figure 5.31.

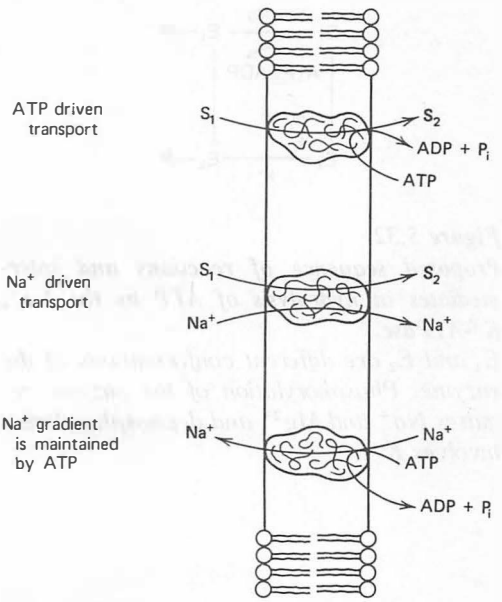


Figure 5.31
Involvement of metabolic energy (ATP) in active mediated transport systems.
The chemical energy released on the hydrolysis of ATP to ADP and inorganic phosphate is used to drive the active transport of various substances, including Na^+. The transmembrane concentration gradient of Na^+ is also used for the active transport of substances.

Active Transport System for Na^+ and K^+: The Na^+, K^+-ATPase

For years a major research effort has been directed toward an explanation of the cellular mechanism for maintenance of the Na^+ and K^+ gradients across the plasma membrane of cells. All mammalian cells contain a Na^+-K^+ antiport system, which utilizes the direct hydrolysis of ATP for movement of ions. Knowledge of this transporter has developed along two paths: (1) from studies of a membrane enzyme which catalyzes ATP hydrolysis and has a requirement for Na^+ and K^+ ions, and (2) from measurements of Na^+ and K^+ movements across intact plasma membranes. It is now accepted that the transporter and the Na^+, K^+-ATPase are one in the same.

All mammalian membranes catalyze the reaction

$$\text{ATP} \xrightarrow[Mg^{2+}]{Na^+ + K^+} \text{ADP} + \text{inorganic phosphate}$$

The enzyme has a requirement for both Na^+ and K^+ ions as well as Mg^{2+}, which is normally required for ATP-requiring reactions. The level of the ATPase in plasma membranes correlates with the Na^+-K^+ transport activity; excitable tissue such as muscle and nerve have a high capacity of both the Na^+-K^+ transport system and the Na^+,K^+-ATPase as do cells actively involved in the movement of Na^+ ion such as the salivary gland and kidney cortex. The protein responsible for the ATPase activity is an oligomer containing two subunits of about 95,000 daltons each and two subunits of about 40,000 with total mol w 270,000. The smallest subunits are glycoproteins, and the complex has the characteristics of a typical integral membrane protein; the enzymatic activity has a requirement for phospholipids indicating its close relationship to membrane function. During the hydrolysis of ATP, the larger subunit is phosphorylated

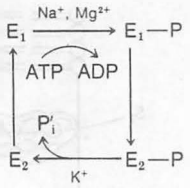


Figure 5.32
Proposed sequence of reactions and intermediates in hydrolysis of ATP by the Na^+, K^+-ATPase.
E_1 and E_2 are different conformations of the enzyme. Phosphorylation of the enzyme requires Na^+ and Mg^{2+} and dephosphorylation involves K^+.

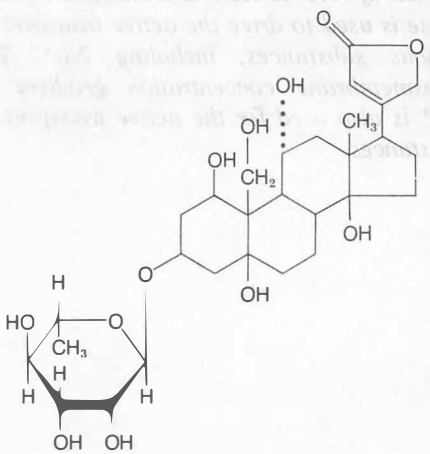


Figure 5.33
Structure of ouabain, a cardiotonic steroid, which is a potent inhibitor of the Na^+, K^+-ATPase and of active Na^+ and K^+ transport.

on a side chain of a specific aspartic acid forming a β-aspartyl phosphate. Phosphorylation of the protein requires Na^+ and Mg^{2+} but not K^+, whereas dephosphorylation of the protein requires K^+ but not Na^+ or Mg^{2+}. The isolated enzyme has an absolute requirement for Na^+, but K^+ can be replaced with NH_4^+ or Rb^+. Several conformations of the protein complex have been observed. A possible sequence of reactions for the enzyme is presented in Figure 5.32.

Of particular significance to its physiological role as a transporter, the enzyme is inhibited by a series of cardiotonic steroids. These pharmacological agents, which include digitalis, increase the force of contraction of heart muscle by altering the excitability of the tissue which is a function of the Na^+-K^+ concentration across the membrane. Ouabain (Figure 5.33) is one of the most active Na^+,K^+-ATPase inhibitors of the series and its action has been studied extensively. The site of binding of ouabain is on the smaller subunit of the enzyme complex and at some distance from the ATP binding site on the larger monomer.

Studies of the transporter activity have been facilitated by use of erythrocyte ghosts, which are intact erythrocyte preparations free of hemoglobin. By carefully adjusting the tonicity of the medium, erythrocytes will swell with breaks in the phospholipid bilayer, permitting the leaking from the cell of cytosolic material, including hemoglobin. The cytosol can be replaced with a defined medium by readjusting the tonicity so that the membrane reseals, trapping the isolation medium inside. In this manner the intracellular ionic and substrate composition and even protein content can be altered. With erythrocyte ghosts the intra- and extracellular Na^+ and K^+ can be manipulated as well as ATP or inhibitor content. With such preparations it has been demonstrated that movement of Na^+ and K^+ is an antiport vectoral process, with Na^+ moving out and K^+ moving into the cell. The ATP binding site on the protein is on the inner surface of the membrane in that hydrolysis occurs only if ATP, Na^+, and Mg^{2+} are inside the cell. K^+ is required externally for dephosphorylation internally of the protein. Ouabain inhibits the translocation of Na^+ and K^+ but only if it is present externally. The protein apparently spans the membrane. The actual number of translocase molecules on an erythrocyte has been estimated by binding studies of radiolabeled ouabain. It is estimated that there are between 100 and 200 molecules per erythrocyte, but the number is significantly larger for other tissues.

ATP hydrolysis by the translocase occurs only if Na^+ and K^+ are translocated, demonstrating that the enzyme is not involved in dissipation of energy in a useless activity. For each ATP hydrolyzed three molecules of Na^+ are moved out of the cell but only two molecules of K^+ in, which leads to an increase in external positive

charges. The electrogenic movement of Na^+ and K^+ is part of the mechanism for the maintenance of membrane potential in a variety of tissues. Even though the energetics of the system dictate that it functions in normal conditions in only one direction, the translocase can be reversed by adjusting the Na^+-K^+ levels; a small net synthesis of ATP has been observed when transport is forced to run in the reverse direction. Obviously, under physiological conditions translocation does not occur in the opposite direction.

A hypothetical model for the movement of Na^+ and K^+ is presented in Figure 5.34. It is proposed that the protein goes through a series of conformational changes during which the Na^+ and K^+ are moved short distances. During the transition a change in the affinity of the binding protein for the cations can occur such that there is a decrease in affinity constants, resulting in the release of the cation into a milieu where the concentration is higher than that from which it was transported.

As an indication of the importance of this enzyme, it has been estimated that the Na^+-K^+ transport system utilizes 60–70% of the ATP synthesized by cells such as nerve and muscle and may utilize about 35% of all ATP generated in a resting individual. A recent report suggests that some obese individuals may have a deficiency in the Na^+-K^+ transport system and thus do not require the same dietary uptake for energy purposes.

Ca^{2+} Translocation in the Sarcoplasmic Reticulum

The sarcoplasmic reticulum of muscle contains an ATP-dependent transport system for Ca^{2+}. Muscle contraction–relaxation cycles are regulated by the cytosolic levels of Ca^{2+} (see Chapter 21, Section 21.1), and the sarcoplasmic reticulum modulates this Ca^{2+} concentration. A Ca^{2+}-dependent ATPase activity has been isolated from the reticulum which has many of the characteristics of the Na^+, K^+-ATPase. The protein apparently spans the membrane and is phosphorylated on an aspartyl residue during the translocation reaction. Two Ca^{2+} ions are translocated for each ATP hydrolyzed. The transport system can move Ca^{2+} against concentration gradients of over 1000 : 1 thereby maintaining very low cytosolic Ca^{2+} levels ($\sim 1 \times 10^{-6}$ M). It has been estimated that the Ca^{2+} translocase may represent 80% of the integral membrane protein of the sarcoplasmic reticulum and occupy a third of the surface area.

Na^+-Dependent Symport Systems

The mechanisms described above for the active transport of cations involved the hydrolysis of ATP as the driving force. Cells have

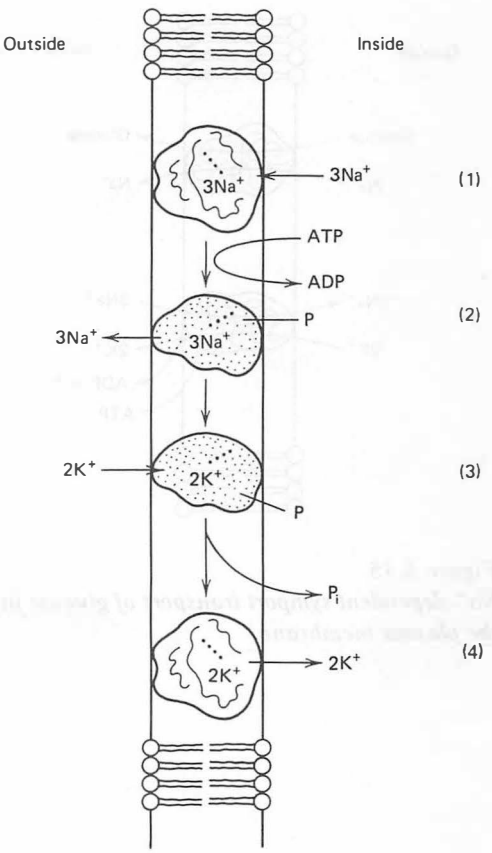


Figure 5.34
A hypothetical model for the translocation of Na^+ and K^+ across the plasma membrane by the Na^+, K^+-ATPase.
(1) Transporter in conformation 1 picks up Na^+. (2) Transporter in conformation 2 translocates and releases Na^+. (3) Transporter in conformation 2 picks up K^+. (4) Transporter in conformation 1 translocates and releases K^+.

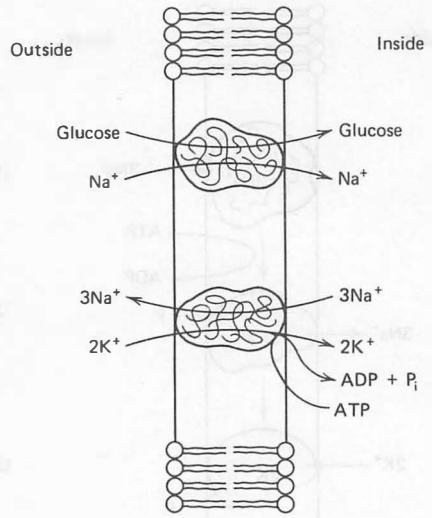


Figure 5.35
Na^+-dependent symport transport of glucose in the plasma membrane.

another energy source in the form of the electrochemical gradient of Na^+ ion which is utilized to actively move sugars and amino acids. A symport translocation system involving the simultaneous movement of both a Na^+ ion and another molecule in the same direction is present in the plasma membrane of cells of the kidney tubule and intestinal epithelium; other tissues may also contain similar transport systems. The general mechanism is presented in Figure 5.35. The diagram represents the transport of D-glucose driven by the movement of Na^+ ion down its concentration gradient. Note that in the transport of the sugar no hydrolysis of ATP occurs. The Na^+-dependent glucose transporter is poorly understood but its activity can be distinguished from passive facilitated diffusion of glucose by differences in inhibitor effects and substrate specificity. There is an absolute requirement for Na^+, and in the process of translocation one Na^+ is moved with each glucose molecule. It can be considered that Na^+ is moving by passive facilitated transport down its electrochemical gradient. It is obligatory that the transporter also translocates a glucose with the Na^+ ion. In the transport the electrochemical gradient of Na^+ ion is dissipated and glucose can be translocated against its concentration gradient. Unless the Na^+ ion gradient is continuously regenerated, the movement of glucose will cease. The Na^+ gradient is maintained by the Na^+-K^+ transport system described above and also represented in Figure 5.35. Thus, metabolic energy in the form of ATP is indirectly involved in glucose transport because it is utilized to maintain the Na^+ ion gradient. Inhibition of energy metabolism and a subsequent decrease in ATP will alter the Na^+ ion gradient and inhibit glucose uptake. Ouabain, the inhibitor of the Na^+-K^+ transporter, inhibits uptake of glucose by preventing the cell from maintaining the Na^+-K^+ gradient. It can be calculated that each glucose molecule requires only one-third of an ATP to be translocated because 3 Na^+ ions are translocated for the hydrolysis of each ATP in the Na^+,K^+-ATPase.

Amino acids are also translocated by the luminal epithelial cells of the intestines by Na^+ ion-dependent pathways similar to the Na^+-dependent glucose transporter. Five different translocases have identified: (1) one for neutral amino acids such as alanine, valine, and leucine; (2) one for basic amino acids, including lysine and arginine; (3) one for the acidic amino acids, aspartate and glutamate; (4) one for the amino acids proline and glycine; and (5) one for amino acids such as β-alanine.

Utilization of the Na^+ gradient as a means to drive the transport of other ions has been reported, including a symport mechanism in the small intestines for the uptake of Cl^- with Na^+ and an antiport mechanism for the secretion of Ca^{2+} out of the cell driven by the simultaneous uptake of Na^+.

It has been proposed that the chemical mechanism for the symport movement of molecules utilizing the Na^+ ion gradient involves a cooperative interaction of the Na^+ ion and the other molecule, that is, glucose, translocated on the protein. A conformational change of the protein occurs following association of the two ligands, which moves the molecules the necessary distance to bring them into contact with the cytosolic environment. The dissociation of the Na^+ ion from the transporter because of the low Na^+ ion concentration inside the cell leads to a return of the protein to its original conformation, a decrease in the affinity for the other ligand, and a release of the ligand into the cytosol.

Group Translocation—A Special Case of Active Transport

As discussed previously, a major hurdle for any active transport system is the release of the transported molecule from the binding site after translocation. If the affinity of the transporter for the translocated molecule did not change there could not be movement against a concentration gradient. In the active transport systems previously described it is believed that there is a change in the affinity for the substance by the transporter by a conformational change of the protein. An alternate mechanism for the release of the substrate is the chemical alteration of the molecule after translocation but before release from the transporter leading to a new compound with a lower affinity for the transporter. The γ-glutamyl cycle for the transport of amino acids across the plasma membrane of some tissues is an example where the substrate is altered during transport and released into the cell as a different molecule. The reactions of the transport mechanism are presented in Figure 5.36. Details of the various chemical reactions in the cycle are presented in Chapter 12. The pathway involves the enzyme γ-glutamyltransferase, which is membrane-bound and catalyzes a transpeptidation reaction, leading to the formation of a dipeptide involving the amino acid transported. The amino acid transported is the substrate to which the γ-glutamyl residue of glutathione (Figure 5.37) is transferred to the amino acid. The new dipeptide is not part of the chemical gradient across the membrane of the original amino acid. The γ-glutamyl derivative is then hydrolyzed by a separate enzyme, not on the membrane, leaving the free amino acid and oxoproline. The process is termed group translocation.

The pathway is active in many tissues but some doubt has been raised about its physiological significance in that individuals have been identified with a genetic absence of the γ-glutamyltransferase activity without any apparent difficulty in amino acid transport. It is

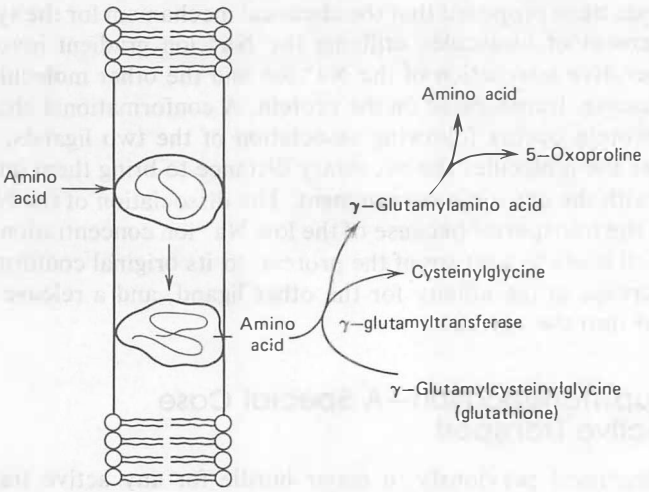


Figure 5.36
The γ-glutamyl cycle.
Represented are the key reactions involved in the group translocation of amino acids across liver cell plasma membranes. The continued uptake of amino acids requires the constant resynthesis of glutathione via a series of ATP requiring reactions described in Chapter 12.

possible, of course, that cells may have several alternate methods for the transport of amino acids and are not dependent on only one mechanism.

All the amino acids except proline can be transported by this group translocation process. The energy for transport comes from the hydrolysis of a peptide bond in glutathione. For the system to continue, glutathione must be resynthesized, which requires the expenditure of three ATP molecules. Thus for each amino acid translocated, three ATPs are required. Recall that the expenditure of only one-third of an ATP is required for each amino acid transported in the Na^+ dependent translocase system. This group translocation is an expensive energetic mechanism for the transport of amino acids.

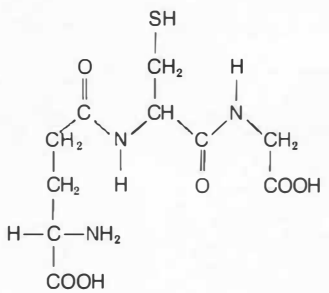


Figure 5.37
Glutathione (γ-glutamylcysteinylglycine).

Summary of Transport Systems

The foregoing has presented the major mechanisms for the movement of molecules across cellular membranes, particularly the plasma membrane. Mitochondria also contain several active transport mechanisms utilizing the pH gradient, that is, a hydrogen ion gradient, developed across the inner membrane. These will be presented in Chapter 6. Bacteria have a number of transport systems analogous to those observed in mammalian cells including passive mediated transporters, one involving a H^+ ion gradient and group translocation.

Table 5.3 summarizes some of the characteristics of the major transport systems found in mammalian cells. (See Clin. Corr. 5.3.)

Table 5.3 Major Transport Systems in Mammalian Cells[a]

Substance Transported	*Mechanism of Transport*	*Tissues*
Sugars		
Glucose	Passive	Most tissues
	Active-symport with Na^+	Small intestines and renal tubular cells
Fructose	Passive	Intestines and liver
Amino acids		
Amino acid specific transporters	Active-symport with Na^+	Intestines, kidney, and liver
All amino acids except proline	Active-group translocation	Liver
Specific amino acids	Passive	Small intestine
Other organic molecules		
Cholic acid, deoxycholic acid, and taurocholic acid	Active-symport with Na^+	Intestines
Organic anions, e.g., malate, α-ketoglutarate, glutamate	Antiport with counterorganic anion	Mitochondria of liver
ATP–ADP	Antiport transport of nucleotides; can be active transport	Mitochondria of liver
Inorganic ions		
Na^+	Passive	Distal renal tubular cells
Na^+/H^+	Active-antiport	Proximal renal tubular cells and small intestines
Na^+/K^+	Active transport-ATP driven	Plasma membrane of all cells
Ca^{2+}	Active-ATP driven	All cells
H^+/K^+	Active transport	Parietal cells of gastric cells secreting H^+
Cl^-/HCO_3^- (perhaps other anions)	Mediated-antiport	Erythrocytes and many other cells

[a] The transport systems are only indicative of the variety of transporters known; others responsible for a variety of substances have been proposed. Most systems have been studied in only a few tissues and their localization may be more extensive than indicated.

CLIN. CORR. 5.3 DISEASES DUE TO LOSS OF MEMBRANE TRANSPORT SYSTEMS

A number of pathological conditions that are due to an alteration in the transport system for specific cellular components have been defined. Some of these are discussed in the appropriate sections describing the metabolism of the intermediates. Individuals have been observed with a decrease in glucose uptake from the intestinal tract due presumably to a loss of the specific glucose transport mechanism. Fructose malabsorption syndromes have been observed, which are also due to an alteration in activity of the facilitated transport system for fructose. In Hartnup's disease there is a decrease in the transport of neutral amino acids in the epithelial cells of the intestines. Little is known concerning possible changes of transport activities in tissues such as muscle, liver, and brain, but it has been suggested that there may be a number of pathological states due to the loss of specific transport mechanisms.

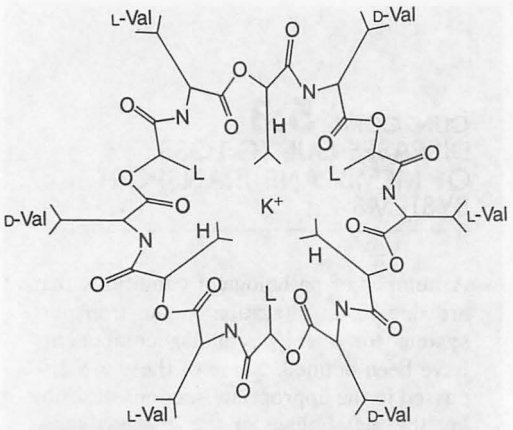


Figure 5.38
The structure of the valinomycin–K^+ complex. *Abbreviations:* D-*Val* = D-*valine;* L-*Val* = L-*valine;* *L* = L-*lactate;* *and* *H* = D-*hydroxyisovalerate.*

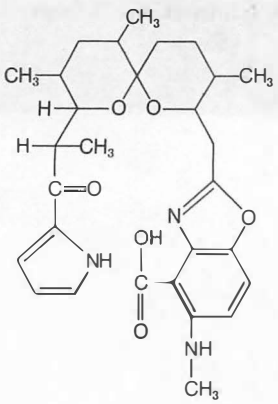


Figure 5.39
Structure of A23187, a Ca^{2+} ionophore.

5.7 IONOPHORES

An interesting class of antibiotics of bacterial origin has been discovered, which facilitates the movement of monovalent and divalent inorganic ions across biological and synthetic lipid membranes. These molecules are not large macromolecules such as proteins but are relatively small molecular weight compounds (up to several thousand); the class of compounds are called ionophores. The possible presence of similar molecules in mammalian tissues has been reported. Ionophores are divided into two major groups: (1) mobile carriers are those ionophores which apparently diffuse back and forth across the membrane carrying the ion from one side of the membrane to the other, and (2) ionophores which apparently form a channel that traverses the membrane and through which ions can diffuse. With both types, molecules are translocated by a passive-mediated transport mechanism. The ionophores which diffuse back and forth across the membranes are more affected by the changes in the fluidity of the membrane than those that form a channel. The major ionophores are listed in Table 5.4.

Each ionophore has a definite ion specificity; valinomycin, whose structure is given in Figure 5.38, has an affinity for K^+ 1,000 times greater than that for Na^+, and the antibiotic A23187 (Figure 5.39) translocates Ca^{2+} 10 times more actively than Mg^{2+}. Several of the diffusion type ionophores that have been studied in detail have a common structural characteristic being cyclic structures, shaped like a doughnut. The metal ion is coordinated to several oxygens in the core of the molecule, and the periphery of the molecule consists of hydrophobic groups. The interaction of the ionophores leads to a chelation of the ion, stripping away its surrounding water shell and encompassing the ion by a hydrophobic shell. The ionophore–ion

Table 5.4 Major Ionophores

Compound	*Major Cations Transported*	*Action*
Valinomycin	K^+ or Rb^+	Uniport, electrogenic
Nonactin	NH_4^+, K^+	Uniport, electrogenic
A 23187	$Ca^{2+}/2H^+$	Antiport, electroneutral
Nigericin	K^+/H^+	Antiport, electroneutral
Monensin	Na^+/H^+	Antiport, electroneutral
Gramicidin	H^+, Na^+, K^+, Rb^+	Forms channels
Alamethecin	K^+, Rb^+	Forms channels

complex is soluble in the lipid membrane and freely diffuses across the membrane. Since the interaction of ion and ionophores is an equilibrium reaction, a steady state develops in the concentration of ions on both sides of the membrane. The specificity of the ionophore is due in part to the size of the pore into which the ion fits and to the attraction of the ionophore for the ion in competition with water molecules.

Valinomycin transports a K^+ by a uniport mechanism and is thus electrogenic, that is, it can create an electrochemical gradient. It carries a positive charge in the form of the K^+ ion across the membrane. Nigericin functions as a antiporter having a free carboxy group, which when dissociated can pick up a K^+ ion, leading to a neutral molecule. Thus on rediffusion back through the membrane it transports a proton; the overall mechanism is electrically neutral, with a K^+ ion exchanging for a H^+ ion. These mechanisms are shown in Figure 5.40.

The other major type of ionophore apparently creates a pore in the membrane. The principal example is gramicidin A. These ionophores show a lower degree of selectivity toward ions, in that the ions are essentially diffusing through a hole in the membrane. Evidence suggests that two molecules of gramicidin A form a channel in the membrane and that the dimer is in constant equilibrium with the free monomer form. By the association and dissociation of the monomers in the membrane, channels can be formed and broken; the rate of interaction of two molecules of gramicidin A would control the rate of ion flux. The structure of the molecules suggests that polar peptide groups line the channel and that hydrophobic groups are on the periphery of the channel interacting with the lipid membrane.

The antibiotic ionophores have been a valuable experimental tool in studies involving ion translocation in biological membranes and for the manipulation of the ionic compositions of cells. There have been reports that proteolipids, prostaglandins, and perhaps other lipids present in mammalian tissues may function as ionophores.

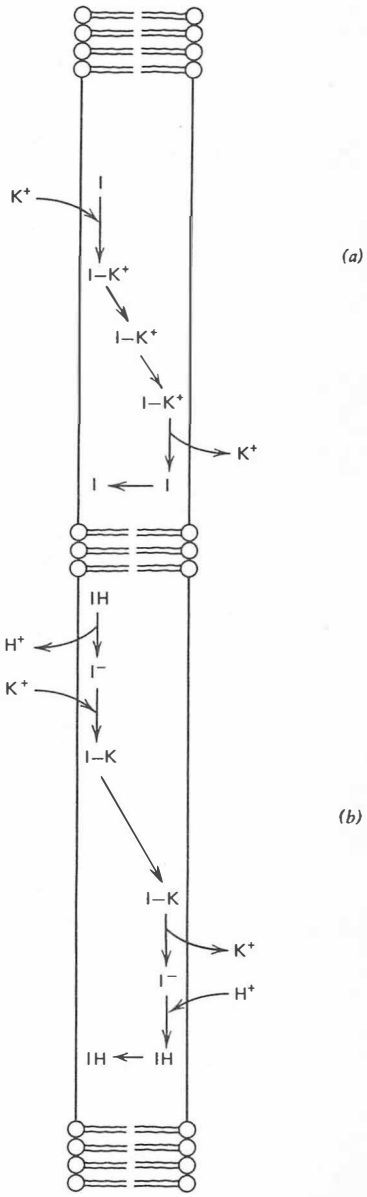


Figure 5.40
Proposed mechanism for the ionophoretic activities of valinomycin and nigericin.
(a) Transport by valinomycin. (b) Transport by nigericin. I represents the ionophore. The valinomycin–K^+ complex is positively charged and translocation of K^+ is electrogenic leading to the creation of a charge separation across the membrane. Nigericin translocates K^+ in exchange for a H^+ across the membrane and the mechanism is electrically neutral.
Diagram adopted from B. C. Pressman, *Annu. Rev. Biochem.*, 45:501, 1976.

BIBLIOGRAPHY

General

Andreoli, T. E., Hoffman, J. F., and Fanestil, D. D. (eds.). *Physiology of membrane disorders*. New York: Plenum, 1978.

Bittar, E. E. (ed.). *Membrane structure and function,* vols. 1–3. New York: Wiley, 1980.

Harrison, R., and Lunt, G. G. *Biological membranes*. New York: Wiley, 1975.

Quinn, P. J. *The molecular biology of cell membranes*. Baltimore: University Park Press, 1976.

Weissmann, G., and Claiborne, R. (eds.). *Cell membranes,* New York: H. P. Publishing Co., 1975.

Membrane structure

Bolis, L., Hoffman, J. F., and Leaf, A. (eds.). *Membranes and disease*. New York: Raven, 1976.

Marchesi, V. T., Furthmayr, H., and Tomita, M. The red cell membrane. *Annu. Rev. Biochem*. 45:667, 1976.

Shinitzky, M., and Henkart, P. Fluidity of cell membranes—current concepts and trends. *Int. Rev. Cyto.* 60:121, 1979.

Singer, S. J., and Nicolson, G. L. The fluid mosaic model of the structure of cell membranes. *Science* 175:720, 1972.

Wallach, D. F. H. *Membrane molecular biology of neoplastic cells*. Amsterdam: Elsevier, 1975.

Transport Processes

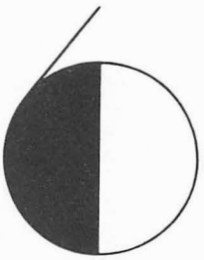
Christensen, H. N. *Biological transport*, 2nd ed. Reading, Mass.: W. A. Benjamin, 1975.

MacLennan, D. H., and Holland, P. C. The Ca^{2+} dependent ATPase of sarcoplasmic reticulum. *Annu. Rev. Biophys. Bioenerg.* 4:377, 1975.

Martonosi, A. (ed.). *The enzymes of biological membranes: membrane transport,* vol. 3, New York: Plenum, 1976.

Skou, J. C., and Norby, D. G. (eds.). *Na^+,K^+-ATPase: structure and kinetics*. New York: Academic, 1979.

Wilson, D. B. Cellular transport mechanisms. *Annu. Rev. Biochem.* 47:933, 1978.



Bioenergetics and Oxidative Metabolism

MERLE S. OLSON

6.1 ENERGY-PRODUCING AND ENERGY-UTILIZING SYSTEMS

Living cells are composed of a complex intricately regulated system of energy-producing and energy-utilizing chemical reactions. Metabolic reactions involved in energy generation sequentially break down ingested or stored macromolecular fuels such as carbohydrate, lipid, or protein in what are termed catabolic pathways. Catabolic reactions usually result in the conversion of large complex molecules to smaller molecules (ultimately CO_2 and H_2O), usually result in the production of storable or conservable energy, and often require the consumption of oxygen during this process. Such reactions are accelerated during periods of fuel deprivation or stress to an organism.

Energy-utilizing reactions are necessary to maintain, to reproduce, or to perform various necessary, and in many instances tissue-specific, cellular functions, for example, nerve impulse conduction and muscle contraction. Metabolic pathways in a cell, which are involved in the biosynthesis of various macromolecules, are termed anabolic pathways. Anabolic reactions usually result in the synthesis of large, complex molecules from smaller precursors and

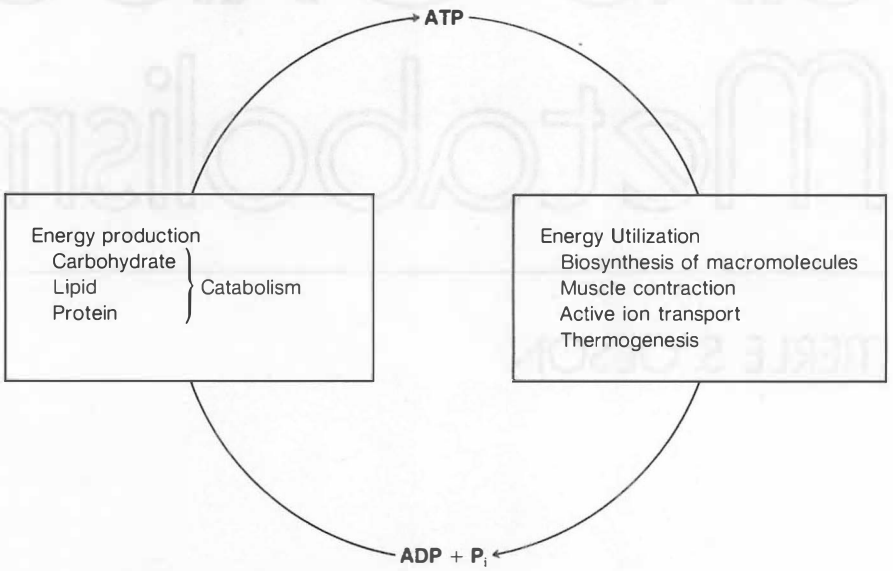
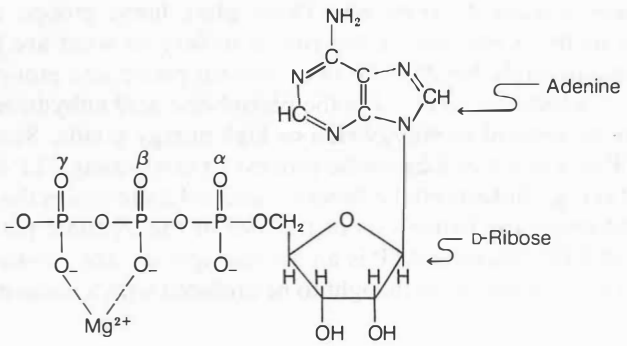


Figure 6.1
The relationship between energy production and energy utilization.

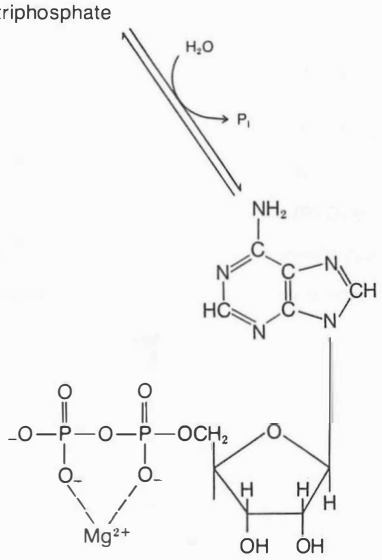
usually require the expenditure of energy. Such reactions are accelerated during periods of relative energy excess, during periods when there occurs a ready availability of precursor molecules or during periods of growth or regeneration of cellular material.

The ATP Cycle

The relationship between the energy-producing and energy-utilizing functions of the cell is illustrated in Figure 6.1. Energy may be derived from the oxidation of appropriate metabolic fuels such as carbohydrate, lipid, or protein. The proportion of each of these metabolic fuels which may be utilized as an energy source depends on the tissue and the dietary and hormonal state of the organism, for example, the mature erythrocyte and the adult brain in the fed



Adenosine 5′-triphosphate



Adenosine 5′-diphosphate

Figure 6.2
Structure of ATP and ADP.

state use only carbohydrate as a metabolic fuel, while the liver of a diabetic or fasted mammal primarily is metabolizing lipid or fat. Energy may be consumed during the performance of various energy-linked (work) functions, some of which are indicated in Figure 6.1. Again the proportion of energy expended or utilized depends largely on the tissue and the physiological state of that tissue, for example, the liver and the pancreas are tissues primarily involved in biosynthetic and secretory work functions, while cardiac and skeletal muscle primarily are involved in converting metabolic energy into mechanical energy during the muscle contraction process.

The essential linkage between the energy-producing and the energy-utilizing pathways is maintained by the nucleoside triphosphate, adenosine 5′-triphosphate (ATP) (Figure 6.2). ATP is a purine (adenine) nucleotide in which the adenine ring is attached in a glycosidic linkage to D-ribose. Three phosphoryl groups are esterified to the 5 position of the ribose moiety in what are termed phosphoanhydride bonds. The two terminal phosphate groups (i.e., β and γ), which are involved in the phosphoric acid anhydride bonding, are designated as energy-rich or high energy bonds. Synthesizing ATP as a result of a catabolic process or consuming ATP in some type of energy-linked cellular function alternately involves the formation and either the hydrolysis or transfer of the terminal phosphate group of ATP. Because ATP is an anionic species, the physiological form of this nucleotide is thought to be chelated with a divalent metal

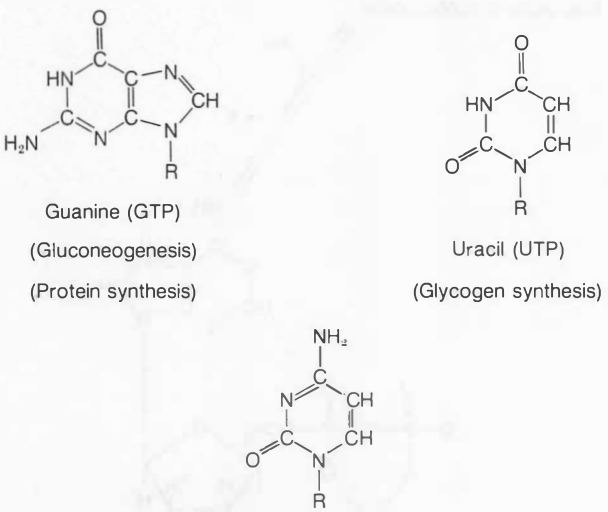


Figure 6.3
Structures of purine and pyrimidine bases involved in various biosynthetic pathways.

cation such as magnesium. ADP also can chelate magnesium, but the affinity of the metal cation for ADP is considerably less than for ATP. While adenine nucleotides are involved intimately in the process of energy generation or conservation, various nucleoside triphosphates, including ATP are involved actively in transferring an energy component into biosynthetic processes. As indicated in Figure 6.3, the guanine nucleotide GTP serves as the source of energy input into the processes of gluconeogenesis and protein synthesis, whereas UTP (uracil) and CTP (cytosine) are utilized in glycogen synthesis and lipid synthesis, respectively. The energy in the terminal phosphate bonds of ATP may be transferred to the other nucleotides, using either the nucleoside diphosphate kinase or the nucleoside monophosphate kinase as illustrated in Figure 6.4. Two nucleoside

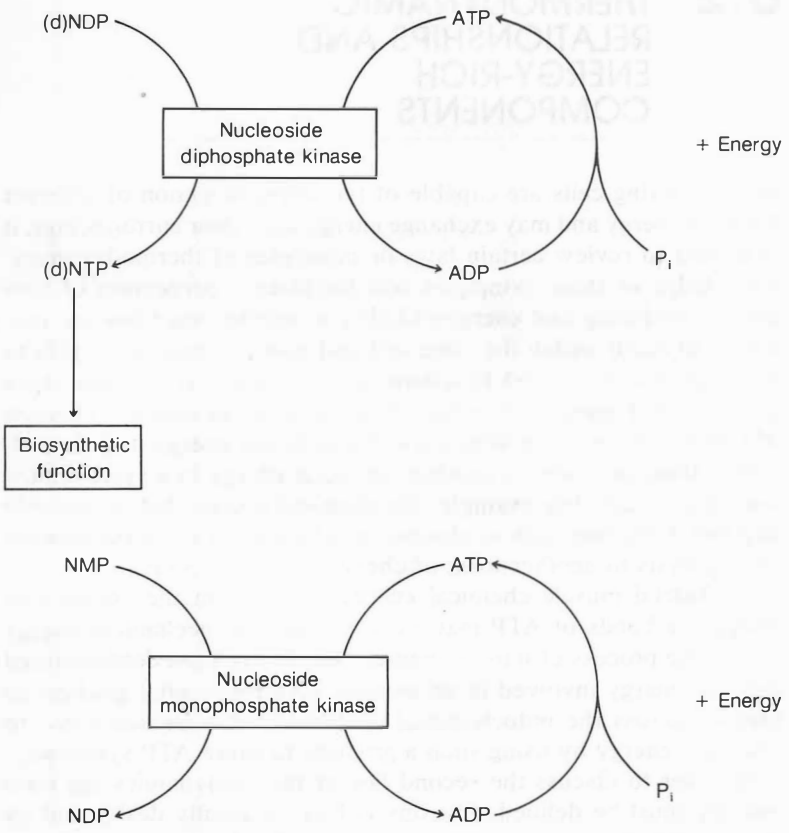


Figure 6.4
Nucleoside diphosphate kinase and nucleoside monophosphate kinase reactions. N = any purine or pyrimidine base.

diphosphates can be converted to a triphosphate and a monophosphate in various nucleoside monophosphate kinase reactions, such as the adenylate kinase reaction

$$2ADP \xleftrightarrow[\text{adenylase kinase (myokinase)}]{} ATP + AMP$$

A consequence of the action of these types of enzymes is that the terminal energy-rich phosphate bonds of ATP may be transferred to the appropriate nucleotides and utilized in a variety of biosynthetic processes.

6.2 THERMODYNAMIC RELATIONSHIPS AND ENERGY-RICH COMPONENTS

Because living cells are capable of the interconversion of different forms of energy and may exchange energy with their surroundings, it is helpful to review certain laws or principles of thermodynamics. Knowledge of these principles will facilitate a perception of how energy-producing and energy-utilizing metabolic reactions are permitted to occur within the same cell and how an organism is able to accomplish various work functions. The first law of thermodynamics indicates that energy can neither be created nor destroyed. This law of energy conservation stipulates that although energy may be converted from one form to another, the total energy in a system must remain constant. For example, the chemical energy that is available in a metabolic fuel such as glucose may be converted in the process of glycolysis to another form of chemical energy, ATP.

In skeletal muscle chemical energy involved in the energy-rich phosphate bonds of ATP may be converted to mechanical energy during the process of muscle contraction. It has been demonstrated that the energy involved in an osmotic electropotential gradient of protons across the mitochondrial membrane may be converted to chemical energy by using such a gradient to drive ATP synthesis.

In order to discuss the second law of thermodynamics the term entropy must be defined. Entropy (which is usually designated by the symbol *S*) is a measure or indicator of the degree of disorder or randomness in a system. Entropy also can be viewed as the energy in a system that is unavailable to perform useful work. All processes,

whether chemical or biological, tend to progress toward a situation of maximum entropy. In fact, unless a process leads to an increase in entropy, it will not occur spontaneously. Equilibrium in a system will result when the randomness or disorder (entropy) is at a maximum. However, it is nearly impossible to quantitate entropy changes in systems that may be useful to study in biochemistry, and such systems are rarely at equilibrium. For the sake of simplicity and its inherent utility in these types of considerations, the quantity termed free energy is employed.

Free Energy

The free energy (denoted by the letter G) of a system is that portion of the total energy in a system that is available for useful work and may be further defined by the equation

$$\Delta G = \Delta H - T\,\Delta S$$

In this expression for a system proceeding toward equilibrium at a constant temperature and pressure, ΔG is, of course, the change in free energy, ΔH is the change in enthalpy or the heat content, T is the absolute temperature, and ΔS is the change in entropy of the system. It can be deduced from this relationship that at equilibrium $\Delta G = 0$. Furthermore, any process that exhibits a negative free energy change will occur spontaneously, energy is given off, and this type of process is called an exergonic reaction. A process that exhibits a positive free energy change will not occur spontaneously; energy from some other source must be applied to this process to allow it to proceed toward equilibrium, and this type of process is termed an endergonic reaction. It should be noted that the change in free energy in a biochemical process is the same regardless of the pathway or mechanism employed to attain the final state. Whereas the rate of a given reaction depends on the free energy of activation, the magnitude of the ΔG is not related to the rate of the reaction. The change in free energy for a chemical reaction is related to the equilibrium constant of that reaction. For example, an enzymatic reaction may be described as

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{C} + \mathrm{D}$$

And an expression for the equilibrium constant may be written as

$$K_{eq} = \frac{[\mathrm{C}][\mathrm{D}]}{[\mathrm{A}][\mathrm{B}]}$$

The free energy change (ΔG) at a constant temperature and pressure is defined as

$$\Delta G = \Delta G^\circ + RT \ln \frac{[C][D]}{[A][B]}$$

where ΔG is the free energy change; ΔG° is the standard free energy change, which is a constant for each individual chemical reaction, where the reactants and products in the reaction are present at concentrations of 1.0 M; R is the gas constant, which is 1.987 cal mol^{-1} K^{-1} or 8.134 J mol^{-1} K^{-1}, depending upon whether the resultant free energy change is expressed in calories or joules per mole; and T is the absolute temperature in Kelvins.

Because at equilibrium $\Delta G = 0$, the expression reduces to

$$\Delta G^\circ = -RT \ln K_{eq}$$

or

$$\Delta G^\circ = -2.3\, RT \log K_{eq}$$

Hence if the equilibrium constant for a reaction can be determined, the standard free energy change (ΔG°) for that reaction also can be calculated. The relationship between ΔG° and K_{eq} is illustrated in Table 6.1. When the equilibrium constant of a reaction is below unity, the reaction is endergonic, and the ΔG° is positive. When the equilibrium constant is above 1, the reaction is exergonic, and the ΔG° is negative.

During any consideration of the energy-producing and energy-utilizing metabolic pathways in cellular systems it is important to understand that the free energy changes characteristic of individual enzymatic reactions in an entire pathway are additive, for example,

$$A \longrightarrow B \longrightarrow C \longrightarrow D$$

$$\Delta G^\circ_{A \to D} = \Delta G^\circ_{A \to B} + \Delta G^\circ_{B \to C} + \Delta G^\circ_{C \to D}$$

Although any given enzymatic reaction in a sequence may have a characteristic positive free energy change, as long as the sum of all the free energy changes is negative, the pathway will proceed spontaneously.

Another way of expressing this principle is that enzymatic reactions with positive free energy changes may be coupled to or driven by reactions with negative free energy changes associated with them. This is an important point because in a metabolic pathway such as the glycolytic pathway various individual reactions either

Table 6.1 Tabulation of Values of K_{eq} and ΔG°

K_{eq}	ΔG° ($kcal \cdot M^{-1}$)
10^{-4}	5.46
10^{-3}	4.09
10^{-2}	2.73
10^{-1}	1.36
1	0
10	−1.36
10^2	−2.73
10^3	−4.09
10^4	−5.46

have positive $\Delta G°$'s or $\Delta G°$'s that are close to 0. On the other hand, there are other reactions that have large and negative $\Delta G°$'s, which drive the entire pathway. The crucial consideration is that the sum of the $\Delta G°$'s for the individual reactions in a pathway must be negative in order for such a metabolic sequence to be thermodynamically feasible.

Caloric Value

During the complete stepwise oxidation of glucose, one of the primary metabolic fuels in cellular systems, a large quantity of energy is available. The free energy released during the oxidation of glucose, whether this oxidation is performed in an instrument used for the combustion of such substances, called a calorimeter, or whether the oxidation occurs in a living functioning cell, is illustrated in the following equation:

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O \qquad \Delta G° = -673{,}000 \text{ cal/mol}$$

When this process is performed under aerobic conditions in most types of cells, there exists a potential to conserve less than half of this "available" energy in the form of ATP. The enzymatic machinery in cellular systems is capable of synthesizing 38 molecules of ATP during the complete oxidation of glucose. The $\Delta G°$'s for the oxidation of other metabolic fuels are listed in Table 6.2. Carbohydrates and proteins (amino acids) have a caloric value of 3–4 kcal/g, while lipid (i.e., the long-chain fatty acid palmitate or the triglyceride tripalmitin) exhibits a caloric value nearly three times greater. The reason that more energy can be derived from lipid than from carbohydrate or protein relates to the average oxidation state of the carbon atoms in these substances.

Table 6.2 The Free Energy Changes and Caloric Values Associated with Various Metabolic Fuels

Compound	*Mol Wt*	$\Delta G°$ *(kcal/mol)*	*Caloric Value (kcal/g)*
Glucose	180	−673	3.74
Lactate	90	−326	3.62
Palmitate	256	−2,380	9.30
Tripalmitin	809	−7,510	9.30
Glycine	75	−234	3.12

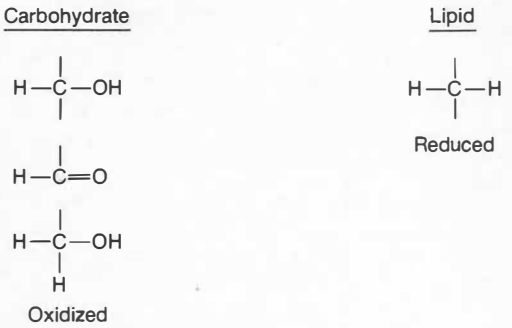


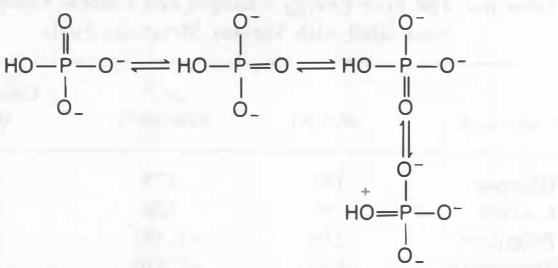
Figure 6.5
Oxidation states of typical carbon atoms in carbohydrates and lipids.

Carbon atoms in carbohydrate are considerably more oxidized (or less reduced) than those in lipid (see Figure 6.5). Hence during the sequential breakdown of these fuels nearly three times as many reducing equivalents (a reducing equivalent is defined as a proton plus an electron, i.e., $H^+ + e^-$) can be extracted from lipid than from carbohydrate. Reducing equivalents may be utilized for ATP synthesis in the mitochondrial energy transduction sequence.

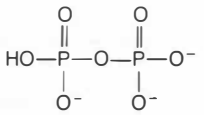
High Energy Compounds

The two terminal phosphoryl groups in the ATP molecule have been identified as energy-rich or high energy bonds. What this description is intended to convey is that the free energy of hydrolysis of such an energy-rich phosphoanhydride bond is much greater than would be obtained for a simple phosphate ester. High energy is not synonymous with stability of the bonding arrangement in question, nor does high energy refer to the energy required to break such bonds. The concept of the high energy compound does imply that the products of the hydrolytic cleavage of the energy-rich bond are in more stable forms than the original compound. As a rule simple phosphate esters (low energy compounds) exhibit negative $\Delta G°$'s of hydrolysis in the range 1–3 kcal/mol, whereas high energy bonds have negative $\Delta G°$'s in the range 5–15 kcal/mol. Simple phosphate esters such as glucose 6-phosphate and glycerol 3-phosphate are examples of low energy compounds. Table 6.3 lists various types of energy-rich compounds with approximate values for their $\Delta G°$'s of hydrolysis.

There are various reasons why certain compounds or bonding arrangements are energy-rich. First, the products of the hydrolysis of an energy-rich bond may exist in more resonance forms than the precursor molecule. The more possible resonance forms in which a molecule can exist tends to stabilize that molecule. The resonance forms for inorganic phosphate can be written as

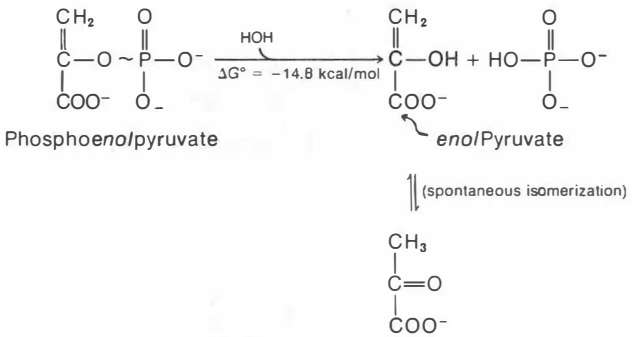


Less resonance forms may be written for ATP or a compound such as pyrophosphate than for inorganic phosphate:



Pyrophosphate

Second, many high energy bonding arrangements have groups of similar electrostatic charge located in close proximity to each other in such compounds. Because like charges tend to repulse one another, the hydrolysis of the energy-rich bond alleviates this situation and, again, lends stability to the products of hydrolysis. Third, the hydrolysis of certain energy-rich bonds results in the formation of an unstable compound, which may isomerize spontaneously to form a more stable compound. The hydrolysis of phospho*enol* pyruvate is an example of this type of compound. The



Pyruvate (Stable form)

$\Delta G°$ of the isomerization reaction is considerable, and the final product, in this case pyruvate, is much more stable. Finally, if a product of the hydrolysis of a high energy bond is an undissociated acid, the dissociation of the proton from the acidic function and its subsequent buffering may contribute to the overall $\Delta G°$ of the hydrolytic reaction. In general, any property or process that lends stability to the products of hydrolysis of a compound tends to confer a high energy character to that compound.

The high energy character of 3′,5′-cyclic AMP has been attributed to the fact that the phosphoanhydride bonding character in this compound is strained as it bridges the 3′ and 5′ positions on the ribose. Further, the energy-rich character of thiol ester compounds such as acetyl CoA or succinyl CoA results from the relatively acidic character of the thiol function. Hence acetyl CoA is nearly equivalent to an anhydride bonding arrangement rather than a simple thioester.

Table 6.3 Examples of Energy-Rich Compounds

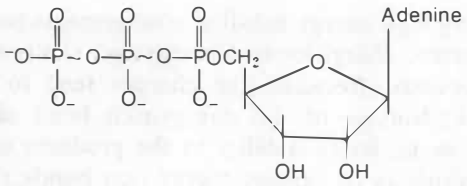
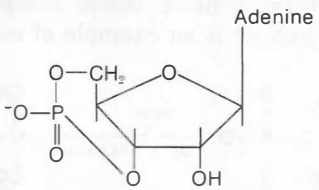
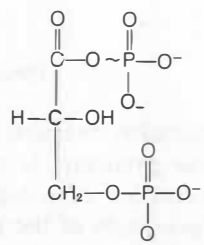
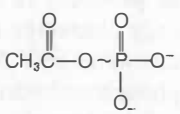
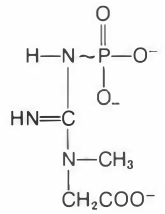
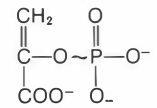
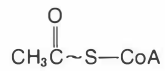
Type of Bond	*$\Delta G°$ of Hydrolysis (kcal/mol)*	*Example*
Phosphoric acid anhydrides	−7.3	 ATP
	−11.9	 3′5′ cyclicAMP
Phosphoric carboxylic acid anhydrides	−10.1	 1,3-Diphosphoglycerate
	−10.3	 Acetyl phosphate

Table 6.3 Examples of Energy-Rich Compounds (continued)

Type of Bond	*$\Delta G°$ of Hydrolysis (kcal/mol)*	*Example*
Phosphoguanidines	−10.3	 Creatine phosphate
Enol phosphates	−14.8	 Phospho*eno*lpyruvate
Thiolesters	−7.7	 Acetyl CoA

Determination of Free Energy Changes

The $\Delta G°$ of hydrolysis of the terminal phosphate of ATP is difficult to determine by simply utilizing the K_{eq} of the hydrolytic reaction because of the position of the equilibrium.

$$ATP + HOH \rightleftharpoons ADP + P_i + H^+$$

However, the $\Delta G°$ of hydrolysis of ATP may be determined indirectly because of the additive nature of free energy changes.

$$\text{Glucose} + \text{ATP} \xrightleftharpoons{\text{hexokinase}} \text{glucose 6-phosphate} + \text{ADP} + \text{H}^+ \qquad \Delta G^\circ = -4.0 \text{ kcal/mol}$$

$$\text{Glucose 6-phosphate} + \text{HOH} \xrightleftharpoons{\text{glucose 6-phosphatase}} \text{glucose} + \text{P}_\text{i} \qquad \Delta G^\circ = -3.3 \text{ kcal/mol}$$

$$\Sigma \quad \text{ATP} + \text{HOH} \rightleftharpoons \text{ADP} + \text{P}_\text{i} + \text{H}^+ \qquad \Delta G^\circ = -7.3 \text{ kcal/mol}$$

Free energy changes for other energy-rich compounds may be determined in a similar fashion.

Transferring High Energy Bond Energies

Energy-rich compounds are capable of transferring various groups from the parent (donor) compound to an acceptor compound in a thermodynamically feasible fashion as long as an appropriate enzyme is present to facilitate the transfer. The energy-rich intermediates in the glycolytic pathway such as 1,3-diphosphoglycerate and phospho*enol*pyruvate can transfer their high energy phosphate moieties to ATP in the phosphoglycerate kinase and pyruvate kinase reactions, respectively.

$$\text{1,3-Diphosphoglycerate} \xrightarrow[\text{ATP}]{\text{phosphoglycerate kinase}} \text{3-phosphoglycerate}$$
(ADP)

$$\text{Phospho}\textit{enol}\text{pyruvate} \xrightarrow[\text{ATP}]{\text{pyruvate kinase}} \text{pyruvate}$$
(ADP)

The ΔG°s of these two reactions are -4.5 and -7.5 kcal/mol, respectively, and hence the transfer of "high energy" phosphate is thermodynamically possible, and ATP synthesis is the result. ATP can transfer its terminal high energy phosphoryl groups to form either compounds of relatively similar high energy character (i.e., creatine phosphate) or compounds that are of considerably lower energy, such as glucose 6-phosphate.

$$\text{Creatine} \xrightleftharpoons[\text{ADP}]{\text{creatine kinase}} \text{creatine phosphate}$$
(ATP)

$$\Delta G^\circ = +1.5 \text{ kcal/mol}$$

$$\text{Glucose} \xrightarrow[\text{ATP}]{\text{hexokinase}} \text{glucose 6-phosphate}$$
(ADP)

$$\Delta G^\circ = -4.0 \text{ kcal/mol}$$

The major point is that phosphate, or for that matter other transferable groups, can be transferred from compounds that contain energy-rich bonding arrangements to compounds that have bonding characteristics of a lower energy in thermodynamically permissible enzymatic reactions. This principle is a major premise of the interaction between energy-producing and energy-utilizing metabolic pathways in living cells.

6.3 SOURCES AND FATES OF ACETYL COENZYME A

Most of the major energy-generating metabolic pathways of the cell eventually result in the production of the two-carbon unit acetyl coenzyme A (CoA). As illustrated in Figure 6.6, the catabolic breakdown of ingested or stored carbohydrate in the glycolytic pathway, long-chain fatty acids resulting from the lipolysis of triglycerides in the β-oxidation sequence or certain amino acids resulting from proteolysis following transamination or deamination and subsequent oxidation provide precursors for the formation of acetyl CoA.

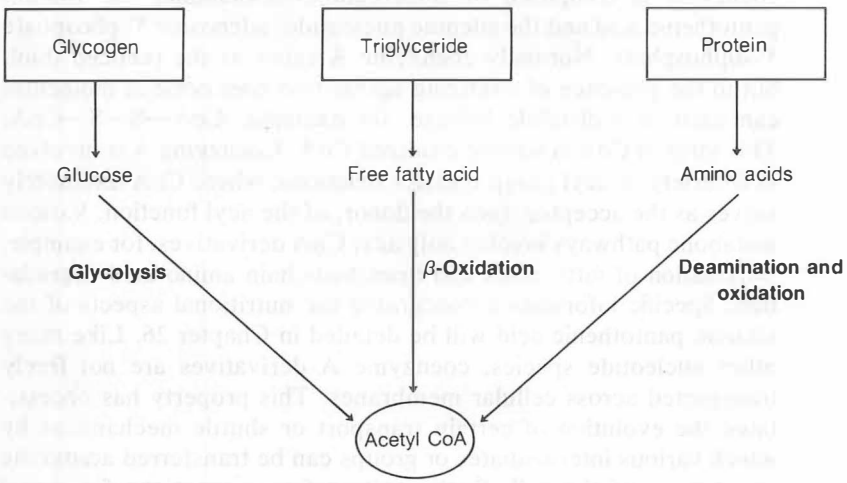


Figure 6.6
General precursors of acetyl CoA.

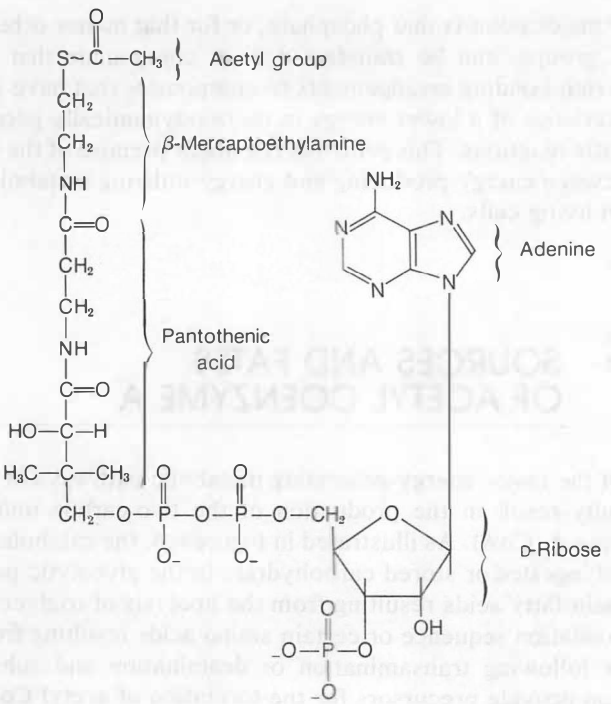


Figure 6.7
Structure of acetyl coenzyme A.

The structure of acetylCoA is shown in Figure 6.7. This complex coenzyme is composed of β-mercaptoethanolamine, the vitamin pantothenic acid and the adenine nucleotide, adenosine 3′-phosphate 5′-diphosphate. Normally coenzyme A exists as the reduced thiol, but in the presence of oxidizing agents two coenzyme A molecules can exist in a disulfide linkage, for example, CoA—S—S—CoA. This form of CoA is termed oxidized CoA. Coenzyme A is involved in a variety of acyl group transfer reactions, where CoA alternately serves as the acceptor, then the donor, of the acyl function. Various metabolic pathways involve only acyl CoA derivatives, for example, β-oxidation of fatty acids and branched-chain amino acid degradation. Specific information concerning the nutritional aspects of the vitamin pantothenic acid will be detailed in Chapter 26. Like many other nucleotide species, coenzyme A derivatives are not freely transported across cellular membranes. This property has necessitated the evolution of certain transport or shuttle mechanisms by which various intermediates or groups can be transferred across the membranes of the cell. Such acyltransferase reactions for acetyl groups and long-chain acyl groups will be discussed in Chapter 9. Finally, as indicated previously, the thiolester linkage in acyl CoA derivatives is an energy-rich bond, and hence these compounds can serve as effective donors of acyl groups in acyltransferase reactions.

Also, in order to synthesize an acylCoA derivative, such as in the acetate thiokinase reaction, a high energy bond of ATP must be expended.

$$\text{Acetate} + \text{CoASH} + \text{ATP} \xrightarrow{\text{acetate kinase}} \text{acetylCoA} + \text{AMP} + \text{PP}_i$$

As was mentioned above, the β-oxidation of fatty acids is a primary source of acetylCoA in many tissues. Whereas a more detailed description of the mobilization, transport, and oxidation of fatty acids is presented in Chapter 9, it is important to note that the products of the β-oxidation sequence are acetyl CoA and reducing equivalents (i.e., NADH + H^+). In certain tissues (e.g., cardiac muscle) and under somewhat special metabolic conditions in other tissues (e.g., in the brain of an individual during prolonged starvation) acetyl CoA for energy generation may be derived from the ketone bodies acetoacetate and β-hydroxybutyrate.

Metabolic Sources and Fates of Pyruvate

During aerobic glycolysis (Chapter 7) glucose or other monosaccharides are converted to pyruvate, and hence in the presence of oxygen pyruvate is the end product per se of this cytosolic pathway. Also the degradation of amino acids such as alanine, serine, and cysteine results in the production of pyruvate (Chapter 12).

Pyruvate has a variety of metabolic fates, depending upon the tissue and the metabolic state of that tissue. The major types of reactions in which pyruvate participates are indicated in Figure 6.8. The oxidative decarboxylation of pyruvate in the pyruvate dehydrogenase reaction is discussed in this chapter; the other reactions in which pyruvate is involved are discussed in Chapter 7.

Pyruvate Dehydrogenase

Pyruvate is converted to acetyl CoA by the pyruvate dehydrogenase multienzyme complex.

$$\text{Pyruvate} + \text{NAD}^+ + \text{CoASH} \xrightarrow{\text{pyruvate dehydrogenase}}$$
$$\text{acetyl CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+ \qquad \Delta G^\circ = -8 \text{ kcal/mol}$$

This enzyme is located exclusively in the mitochondrial compartment and is present in high concentrations in tissues such as cardiac muscle and kidney. Because of the large negative ΔG° of this reaction, under physiological conditions, the pyruvate dehydrogenase reaction is essentially irreversible, and this fact is the

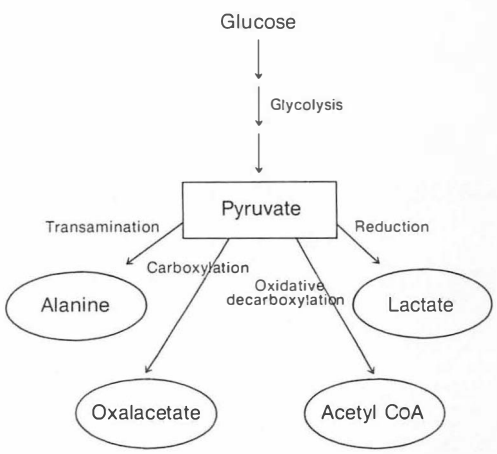


Figure 6.8
Metabolic fates of pyruvate.

primary reason that a net conversion of fatty acid carbon to carbohydrate cannot occur, for example,

$$\text{Acetyl CoA} \not\rightarrow \text{pyruvate}$$

Molecular weights of the multienzyme complex derived from kidney, heart, or liver range from 7 to 8.5×10^6. The mammalian pyruvate dehydrogenase enzyme complex consists of three different types of catalytic subunits:

Number of Subunits/ Complex	*Type*	*Molecular Weight*	*Subunit Structure*
20 or 30[a]	Pyruvate dehydrogenase	154,000	$\alpha_2\beta_2$ Tetramer
60	Dihydrolipoyl transacetylase	52,000	Identical
5–6	Dihydrolipoyl dehydrogenase	110,000	α_2 Dimer

[a] Depending on source.

The structure of the pyruvate dehydrogenase complex derived from *E. coli* (particle weight, 4.6×10^6) is somewhat different from the

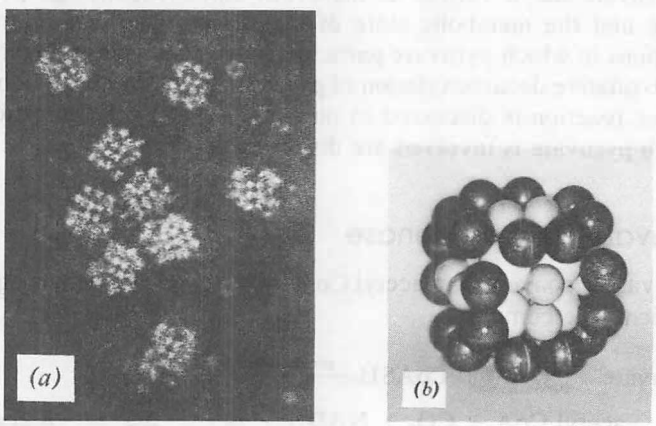


Figure 6.9
***The pyruvate dehydrogenase from* E. coli.**
(a) Electron micrograph. (b) Molecular model. The enzyme complex was negatively stained with phosphotungstate. (X200,000) Courtesy of Dr. Lester J. Reed, University of Texas, Austin.

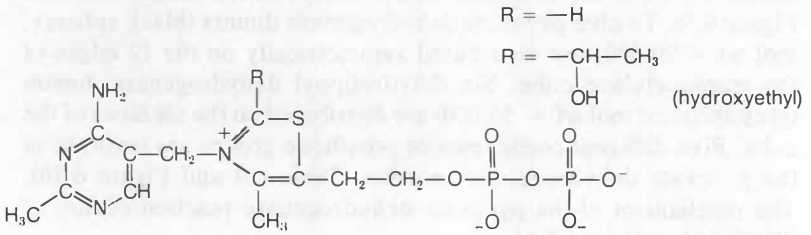
mammalian enzyme. Electron micrographs of the bacterial enzyme complex (Figure 6.9) indicate that the transacetylase, which consists of 24 identical polypeptide chains (mol wt = 64,500), forms the cubelike core of the complex (white spheres in the model shown in Figure 6.9). Twelve pyruvate dehydrogenase dimers (black spheres; mol wt = 90,500) are distributed symmetrically on the 12 edges of the transacetylase cube. Six dihydrolipoyl dehydrogenase dimers (grey spheres; mol wt = 56,000) are distributed on the six faces of the cube. Five different coenzymes or prosthetic groups are involved in the pyruvate dehydrogenase reaction (Table 6.4 and Figure 6.10). The mechanism of the pyruvate dehydrogenase reaction occurs as illustrated in Figure 6.11.

Because of the active participation of thiol groups in the catalytic mechanism of the enzyme, agents which either oxidize or complex with thiol groups are strong inhibitors of the enzyme complex. Arsenite is an example of such an inhibitor.

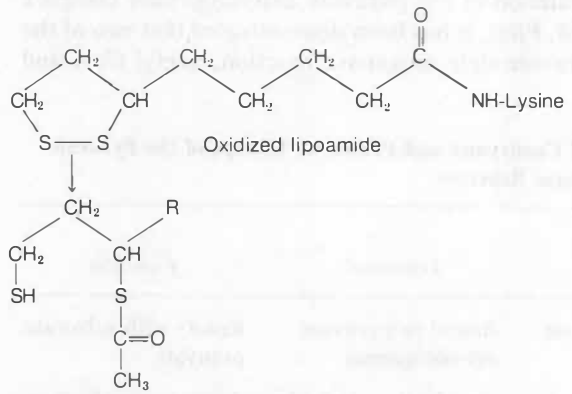
Two types of regulation of the pyruvate dehydrogenase complex have been elucidated. First, it has been demonstrated that two of the products of the pyruvate dehydrogenase reaction, acetyl CoA and

Table 6.4 Function of Coenzymes and Prosthetic Groups of the Pyruvate Dehydrogenase Reaction

Coenzyme or Prosthetic Group	*Location*	*Function*
Thiamin pyrophosphate	Bound to pyruvate dehydrogenase	Reacts with substrate, pyruvate
Lipoic acid	Covalently attached to a lysine residue on the dihydrolipoyl transacetylase	Accepts acetyl group from thiamine pyrophosphate
Coenzyme A	Free in solution	Accepts acetyl group from lipoamide group on the transacetylase
Flavin adenine dinucleotide (FAD)	Tightly bound to dihydrolipoyl dehydrogenase	Accepts reducing equivalents from reduced lipoamide group
Nicotinamide adenine dinucleotide	Free in solution	Terminal acceptor of reducing equivalents from the reduced flavoprotein



Thiamin pyrophosphate



Oxidized lipoamide

Reduced/acetylated lipoamide

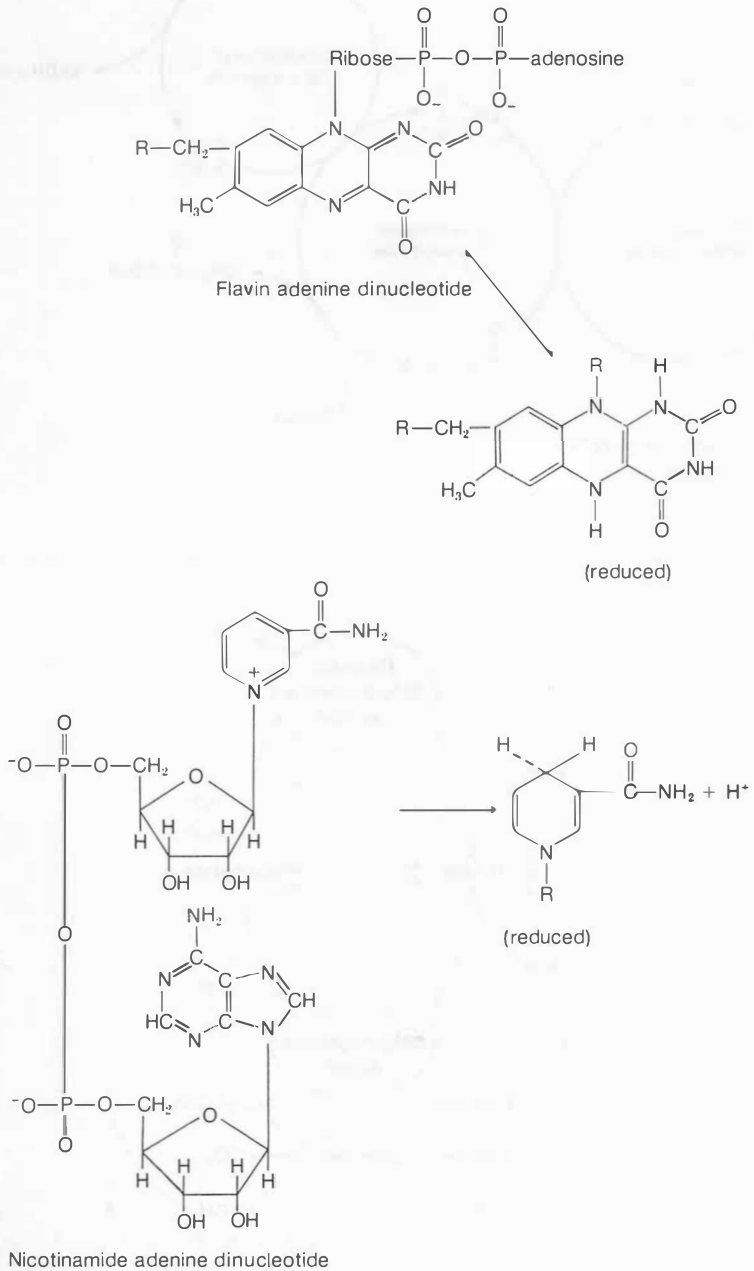
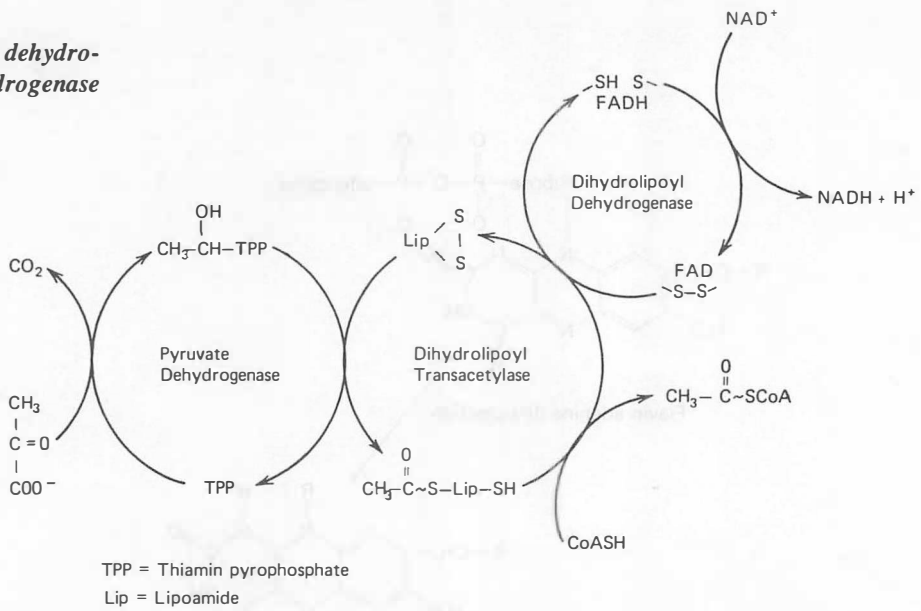


Figure 6.10
Structures of the coenzymes involved in the pyruvate dehydrogenase reaction.
See also Figure 6.7.

Figure 6.11
The mechanism of the pyruvate dehydrogenase reaction; the pyruvate dehydrogenase multienzyme complex.



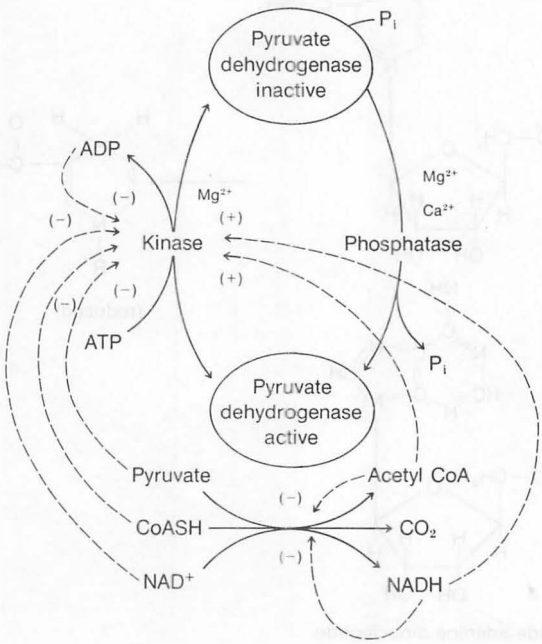


Figure 6.12
The regulation of the pyruvate dehydrogenase multienzyme complex.

NADH, inhibit the complex in a competitive fashion. Second, the pyruvate dehydrogenase complex exists in two forms: (1) an active enzyme complex, and (2) an inactive, phosphorylated complex. The inactivation of the complex is accomplished by a $Mg–ATP^{2+}$-dependent protein kinase, which is tightly bound to the enzyme complex. The reactivation of the complex is accomplished by a phosphoprotein phosphatase, which dephosphorylates the complex in a Mg^{2+}- and Ca^{2+}-dependent reaction. Three separate serine residues on the α subunit of the pyruvate dehydrogenase are phosphorylated by the protein kinase, but the phosphorylation of only one of these sites is related to the activity of the complex. The differential regulation of the pyruvate dehydrogenase kinase and phosphatase is the key to the regulation of pyruvate dehydrogenase complex. The essential features of this complex regulatory system are illustrated in Figure 6.12. Not only can acetyl CoA and NADH, the products of the pyruvate dehydrogenase reaction, inhibit the dephospho- (active) form of the enzyme, but these two compounds stimulate the protein kinase reaction, leading to an interconversion of the complex to its inactive form. In addition, free CoASH and NAD^+ inhibit the protein kinase. Hence with any increase of the mitochondrial NADH/NAD^+ or acetyl CoA/CoA ratio, such as during rapid β-oxidation of fatty acids, pyruvate dehydrogenase will be inactivated by the kinase reaction. Also the substrate of the enzyme complex, pyruvate, is a potent inhibitor of the protein kinase, and therefore in the presence of elevated tissue pyruvate levels the kinase will be inhibited and the complex maximally active. Finally, it has been demonstrated that insulin administration can activate pyruvate dehydrogenase in adipose tissue, and catecholamines, that is, epinephrine, can activate pyruvate dehydrogenase in cardiac tissue. The mechanisms of these hormonal effects are not well understood, but alterations of the intracellular distribution of calcium, such that the phosphoprotein phosphatase reaction is stimulated in the mitochondrial compartment, may be involved in these effects. These hormonal effects are not mediated directly by alterations in the tissue cAMP levels because the pyruvate dehydrogenase protein kinase and phosphatase are cAMP-independent or insensitive. (See Clin. Corr. 6.1.)

Metabolic Fates of Acetyl CoA

The various fates of acetyl CoA generated in the mitochondrial compartment include (1) complete oxidation of the acetyl group in the tricarboxylic acid cycle for energy generation; (2) in a tissue such as the liver, conversion of an excess of acetyl CoA into the ketone bodies, acetoacetate and β-hydroxybutyrate; and (3) a transfer of

CLIN. CORR. 6.1 PYRUVATE DEHYDROGENASE DEFICIENCY

A variety of disorders in pyruvate metabolism has been detected in children. Some of these defects have been shown to involve deficiencies in each of the different component catalytic or regulatory subunits of the pyruvate dehydrogenase multienzyme complex. Children detected with a pyruvate dehydrogenase deficiency usually exhibit elevated serum levels of lactate, pyruvate, and alanine, which produce a chronic lactic acidosis. Such patients frequently exhibit severe neurological defects, and in most situations this type of enzymatic defect results in death. The diagnosis of pyruvate dehydrogenase deficient patients is usually made by assaying the enzyme complex and/or its various enzymatic subunits in cultures of skin fibroblasts taken from the patient. In certain instances patients respond to dietary management in which a ketogenic diet is administered and carbohydrates are minimized.

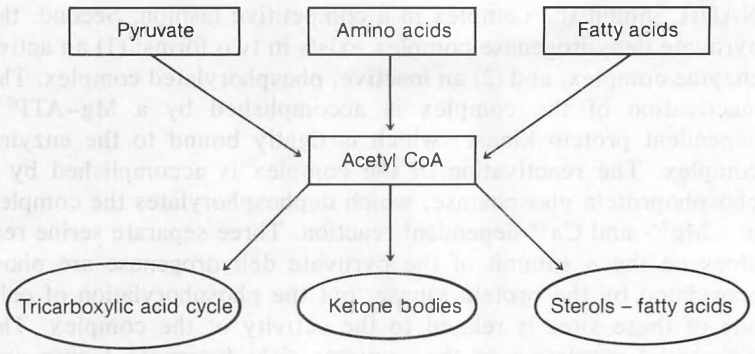


Figure 6.13
Sources and fates of acetyl CoA.

the acetyl units to the cytosol with subsequent biosynthesis of such complex molecules as sterols (Chapter 10) and long-chain fatty acids (Chapter 9) (see Figure 6.13).

6.4 THE TRICARBOXYLIC ACID CYCLE

The primary metabolic fate of acetyl CoA produced in the various energy-generating catabolic pathways of most cells is its complete oxidation in a cyclic series of oxidative reactions termed the *tricarboxylic acid cycle*. This metabolic cycle also is commonly referred to as the citric acid cycle or the Krebs cycle after Sir Hans Krebs who postulated the essential features of this pathway in 1937. Various investigators defined many of the enzymes and di- and tricarboxylic acid intermediates in this pathway, but it was Krebs who pieced together these components in his formulation of the "Krebs cycle." Although certain of the cycle enzymes are found in the cytosol, the primary location of enzymes of the tricarboxylic acid cycle is in the mitochondrion. This type of distribution is appropriate because the pyruvate dehydrogenase multienzyme complex and the fatty acid β-oxidation sequence, the two primary sources for generating acetyl CoA, are located in the mitochondrial compartment. Also, one of the primary functions of the tricarboxylic acid cycle is to generate reducing equivalents, which are utilized to gen-

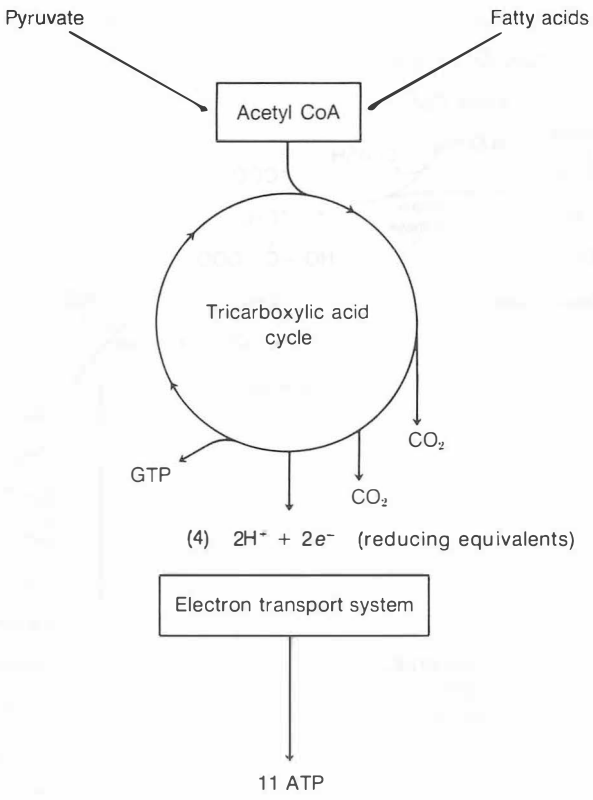


Figure 6.14
General description of mitochondrial ATP synthesis.

erate energy, that is, ATP, in the electron transport–oxidative phosphorylation sequence, another process contained exclusively in the mitochondrion (see Figure 6.14). Mitochondrial energy transduction is discussed in Section 6.7.

The individual enzymatic reactions of the tricarboxylic acid cycle are illustrated in Figure 6.15. Figure 6.14 illustrates the essential process involved in the Krebs cycle. The substrate or input into the cycle is the two-carbon unit acetyl CoA, and the products of a complete turn of the cycle are $2CO_2$ plus one high energy phosphate bond (as GTP) and four reducing equivalents (i.e., 3NADH + H^+ and 1$FADH_2$).

Individual Reactions of the Tricarboxylic Acid Cycle

The initial step of the cycle is catalyzed by the enzyme citrate synthase. This is a highly exergonic reaction and essentially commits acetyl groups toward citrate formation or oxidation in the Krebs

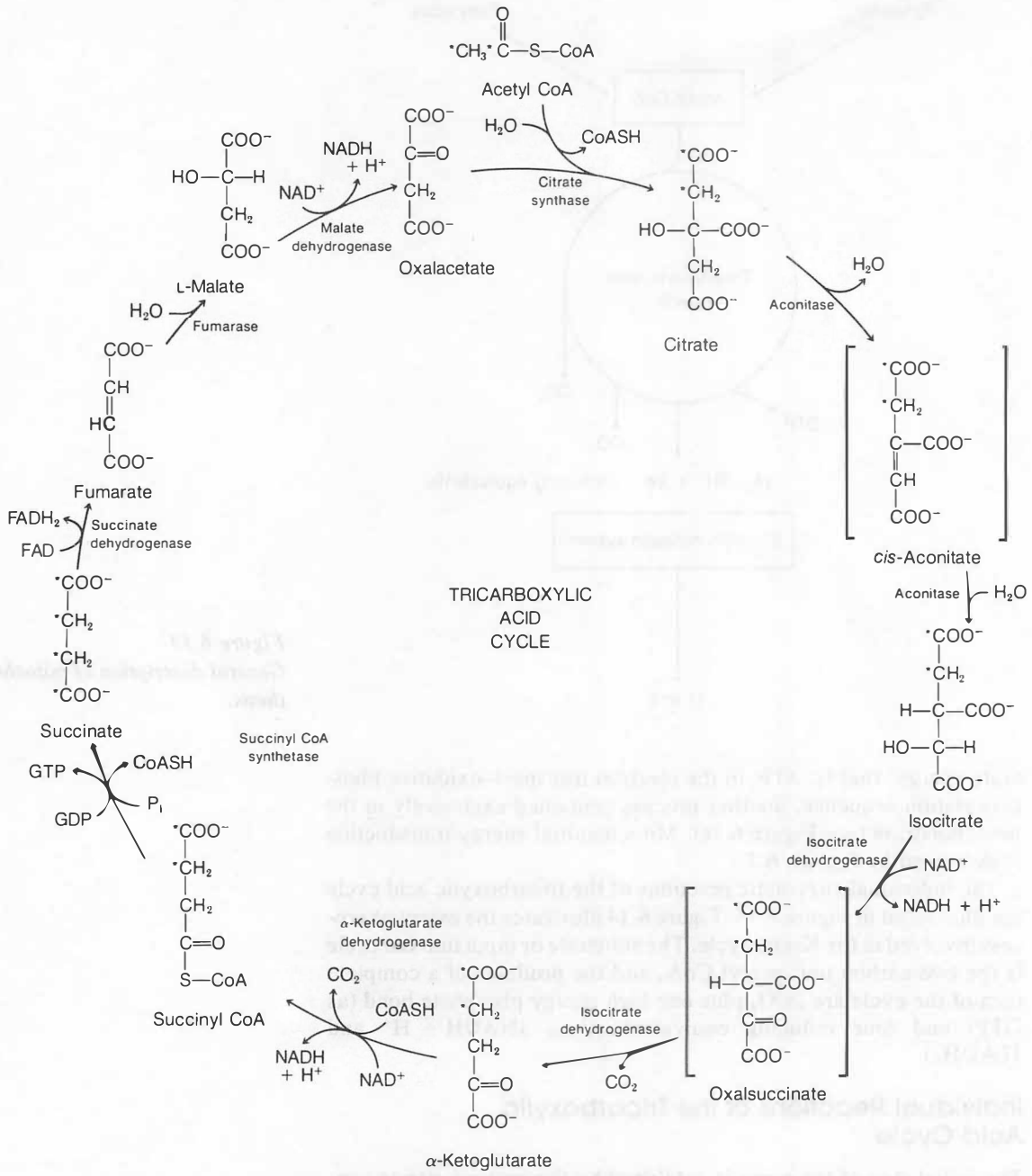
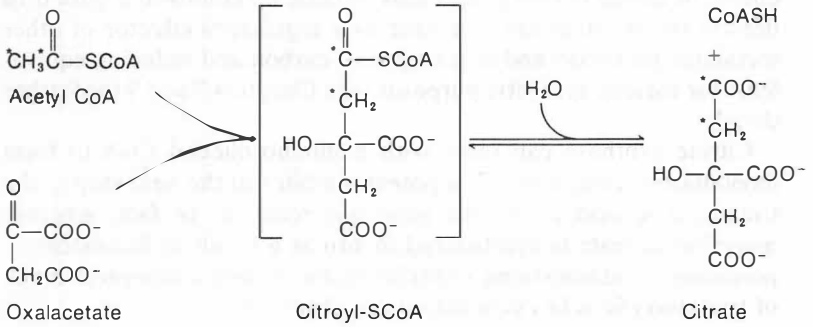


Figure 6.15
The tricarboxylic acid cycle.

cycle. As shown below the citrate synthase reaction involves the condensation of the acetyl moiety and the α-keto function of the dicarboxylic acid oxalacetate. Citrate synthase is an enzyme with mol wt 100,000 and exists in the mitochondrial matrix.



The equilibrium of this reaction is far toward citrate formation with a $\Delta G°$ near -9 kcal/mol. The citroyl-SCoA intermediate in this reaction is not released from the enzyme during the reaction and is thought to remain bound to the catalytic site on citrate synthase. It has been estimated that the citrate synthase reaction is considerably displaced from equilibrium under in situ conditions, which makes this step a primary candidate for regulatory modulation. It has been proposed that this enzyme is regulated (inhibited) by ATP, NADH, succinyl CoA, and long-chain acyl CoA derivatives on the basis of experiments performed with the purified enzyme, but none of these

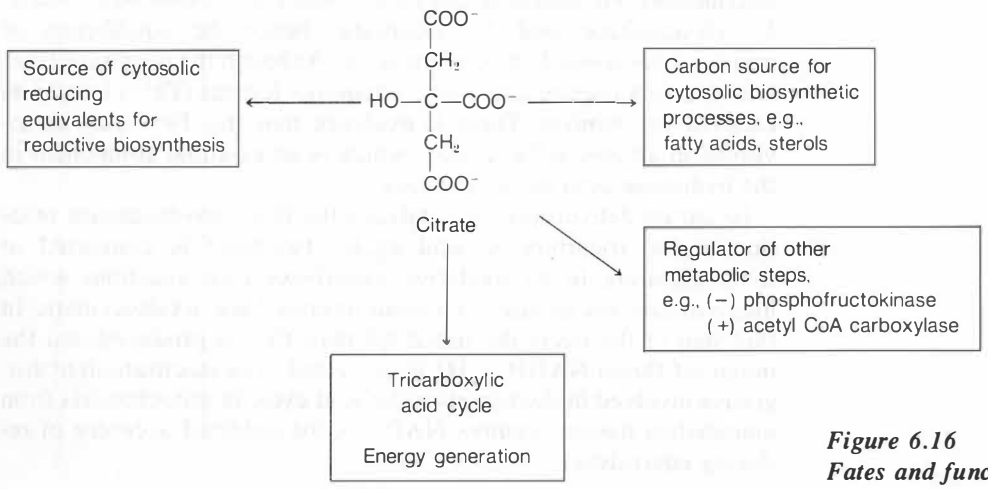


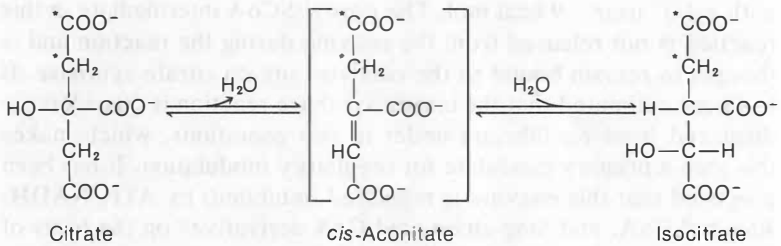
Figure 6.16
Fates and functions of citrate.

effects has been proven to be operative in intact metabolic systems under physiological conditions.

It is most probable that the primary regulator of the citrate synthase reaction is the availability of its two substrates, acetyl CoA and oxalacetate. It is important to note the many important fates and effects of citrate in energy and biosynthetic metabolism. Figure 6.16 depicts the involvement of citrate as a regulatory effector of other metabolic pathways and as a source of carbon and reducing equivalents for various synthetic purposes (see Chapters 7 and 9 for further details).

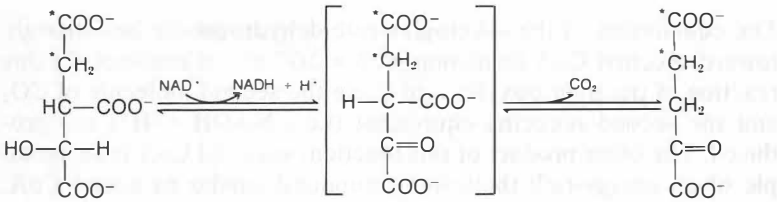
Citrate synthase can react with monofluoroacetyl CoA to form monofluorocitrate, which is a potent inhibitor of the next step in the tricarboxylic acid cycle, the aconitase reaction. In fact, whether monofluorocitrate is synthesized in situ as a result of fluoroacetate poisoning or administered experimentally, nearly a complete block of tricarboxylic acid cycle activity is observed.

Citrate is converted to isocitrate in the aconitase reaction.



Again, this reaction involves the generation of an enzyme-bound intermediate, *cis*-aconitate. At equilibrium there exists 90% citrate, 3% *cis*-aconitate, and 7% isocitrate, hence the equilibrium of aconitase lies toward citrate formation. Although the aconitase reaction does not require cofactors, it requires ferrous (Fe^{2+}) iron in its catalytic mechanism. There is evidence that this Fe^{2+} may be involved in an iron-sulfur center, which is an essential component in the hydratase activity of aconitase.

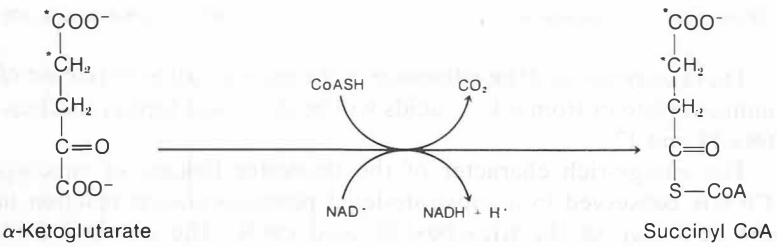
Isocitrate dehydrogenase catalyzes the first dehydrogenase reaction in the tricarboxylic acid cycle. Isocitrate is converted to α-ketoglutarate in an oxidative decarboxylation reaction, which likely occurs via an enzyme-bound intermediate, oxalsuccinate. In this step of the cycle the initial (of two) CO_2 is produced and the initial (of three) NADH + H^+ is generated. The isocitrate dehydrogenase involved in the tricarboxylic acid cycle in mitochondria from mammalian tissues requires NAD^+ as the oxidized acceptor of reducing equivalents.



Mitochondria also possess an isocitrate dehydrogenase that requires $NADP^+$ as the oxidized coenzyme. The $NADP^+$-linked enzyme also may be found in the cytosol, where it is probable that it is involved in providing reducing equivalents for cytosolic reductive processes. The equilibrium of this reaction lies strongly toward α-ketoglutarate formation with a $\Delta G°$ of nearly -5 kcal/mol. The NAD^+-linked isocitrate dehydrogenase has mol wt 380,000 and consists of eight identical subunits. The reaction requires a divalent metal cation (e.g., Mn^{2+} or Mg^{2+}) in the decarboxylation of the β position of the oxalsuccinate. The NAD^+-linked isocitrate dehydrogenase is stimulated by ADP and in some cases AMP and is inhibited by ATP and NADH. Hence under high energy conditions (i.e., high ATP/ADP + P_i and high NADH/NAD^+ ratios) the NAD^+-linked isocitrate dehydrogenase of the tricarboxylic acid cycle is inhibited. During periods of low energy, on the other hand, the activity of this enzyme is stimulated in order to accelerate energy generation in the tricarboxylic acid cycle.

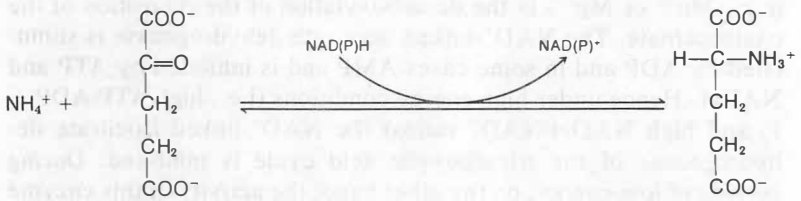
The conversion of α-ketoglutarate to succinyl CoA is catalyzed by the α-ketoglutarate dehydrogenase multienzyme complex. This enzyme complex is nearly identical to the pyruvate dehydrogenase in terms of the reactions catalyzed and some of the structural features of the complex. Again, thiamine pyrophosphate, lipoic acid, CoASH, FAD, and NAD^+ participate in the catalytic mechanism of the α-ketoglutarate dehydrogenase reaction.

The multienzyme complex consists of the α-ketoglutarate dehydrogenase, the dihydrolipoyl transuccinylase and the dihydrolipoyl dehydrogenase as the three catalytic subunits.



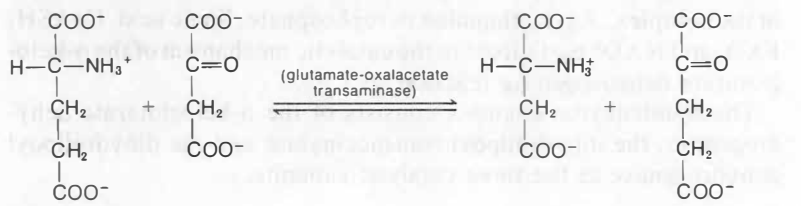
The equilibrium of the α-ketoglutarate dehydrogenase lies strongly toward succinyl CoA formation with a ΔG° of −8 kcal/mol. In this reaction of the tricarboxylic acid cycle the second molecule of CO_2 and the second reducing equivalent (i.e., NADH + H^+) are produced. The other product of this reaction, succinyl CoA is an example of an energy-rich thiolester compound similar to acetyl CoA. Unlike the pyruvate dehydrogenase complex the α-ketoglutarate dehydrogenase is not regulated by a protein kinase-mediated phosphorylation reaction. The nucleoside triphosphates, ATP and GTP, NADH, succinyl CoA and Ca^{2+} have been shown to inhibit this enzyme complex.

It is at the level of α-ketoglutarate in the Krebs cycle where intermediates may leave this oxidative pathway to be reductively aminated in the glutamate dehydrogenase reaction. This mitochondrial enzyme converts α-ketoglutarate to glutamate in the presence of NADH or NADPH and ammonia.



α-Ketoglutarate Glutamate

Using various transamination reactions the amino group thus incorporated into glutamate is transferred to a variety of other amino acids.

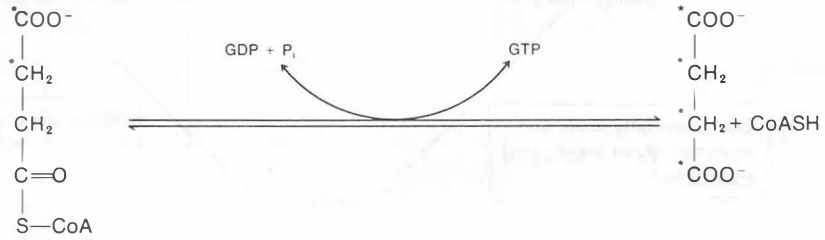


Glutamate Oxalacetate Aspartate α-Ketoglutarate

These enzymes and the relevance of the incorporation or release of ammonia into or from α-keto acids will be discussed further in Chapters 11 and 12.

The energy-rich character of the thiolester linkage of succinyl CoA is conserved in a substrate-level phosphorylation reaction in the next step of the tricarboxylic acid cycle. The succinyl CoA synthetase or succinate thiokinase reaction converts succinyl CoA to

succinate and in mammalian tissue results in the phosphorylation of GDP to GTP.



Succinyl CoA Succinate

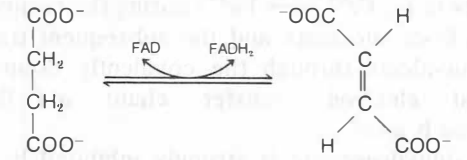
This reaction is freely reversible with a $\Delta G^\circ = -0.7$ kcal/mol and the catalytic mechanism involves an enzyme–succinyl phosphate intermediate:

Succinyl CoA + P_i + Enz ⟷ Enz—succinyl phosphate + CoASH
Enz—succinyl phosphate ⟷ Enz—phosphate + succinate
Enz—phosphate + GDP ⟷ Enz + GTP

The enzyme is phosphorylated on the 3 position of a histidine residue during the succinyl CoA synthetase reaction. Hence in this step of the tricarboxylic acid cycle a high energy bond is conserved as GTP. Because of the presence of the nucleoside diphosphokinase discussed earlier in this chapter, this GTP may be converted to ATP.

Succinyl CoA represents a metabolic branch point in that intermediates may enter or exit the Krebs cycle at this point (see Figure 6.17). Succinyl CoA may be formed either from α-ketoglutarate in the Krebs cycle or from methylmalonyl CoA in the final steps of the breakdown of odd chain length fatty acids or the branched-chain amino acids valine and isoleucine. The metabolic fates of succinyl CoA include its conversion to succinate in the succinyl CoA synthetase reaction of the Krebs cycle and its condensation with glycine to form δ-aminolevulinate in the δ-aminolevulinate synthetase reaction, which is the initial reaction in porphyrin biosynthesis (see Chapter 22).

Succinate is oxidized to fumarate in the succinate dehydrogenase reaction of the Krebs cycle.



Succinate Fumarate

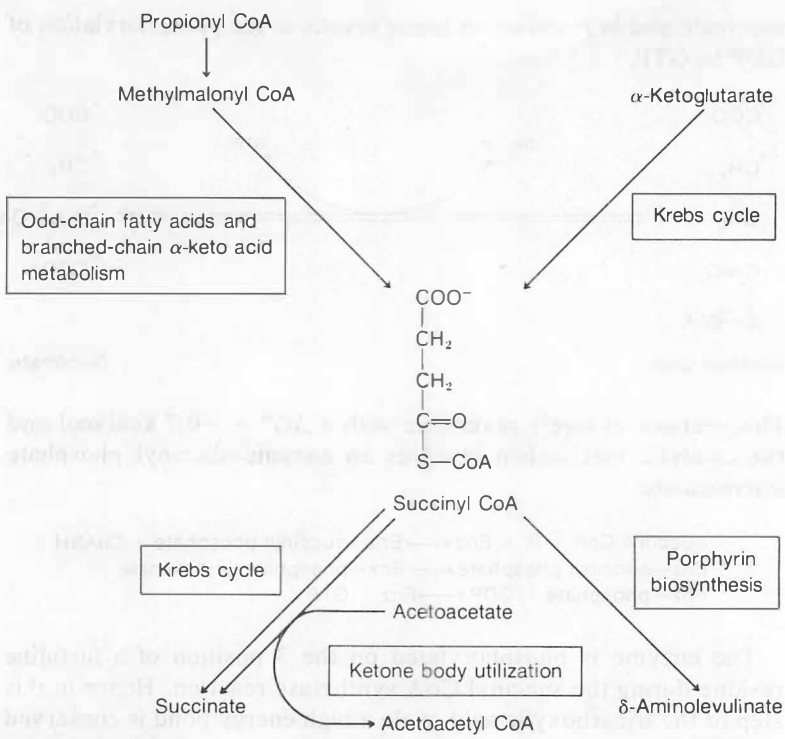
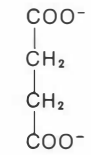


Figure 6.17
Sources and fates of succinyl CoA.

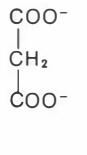
Succinate dehydrogenase is tightly bound to the inner mitochondrial membrane and is composed to two subunits with mol w 70,000 and 30,000. The 70,000 mol w subunit contains the substrate binding site, the covalently bound (to a lysine residue) FAD, 4 nonheme iron atoms, and 4 acid-labile sulfur atoms, while the 30,000 mol w subunit contains 4 nonheme irons and 4 acid-labile sulfur atoms. It is thought that this enzyme is a typical example of an iron-sulfur protein in which the nonheme iron of succinate dehydrogenase undergoes valence changes (e.g., $Fe^{2+} \longrightarrow Fe^{3+}$) during the removal of electrons and protons from succinate and the subsequent transfer of these reducing equivalents through the covalently bound FAD to the mitochondrial electron transfer chain at the coenzyme Q–cytochrome b level.

Succinate dehydrogenase is strongly inhibited by malonate and

oxalacetate and is activated by ATP, inorganic phosphate, and succinate. Malonate inhibits succinate dehydrogenase competitively with respect to succinate. This inhibitory characteristic of malonate is due to the very close structural similarity between malonate and the substrate succinate. Malonate is used as a very effective inhibitor of the Krebs cycle in complex metabolic systems. In fact, the ability of malonate to inhibit the cycle at the succinate dehydrogenase step was used by Krebs as evidence for the cyclic nature of this oxidative metabolic pathway.

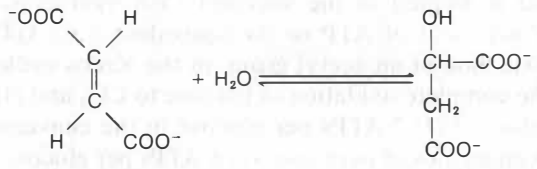


Succinate



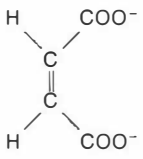
Malonate

Fumarate is hydrated to form L-malate in the next step in the tricarboxylic acid cycle by the enzyme fumarase. Fumarase is a



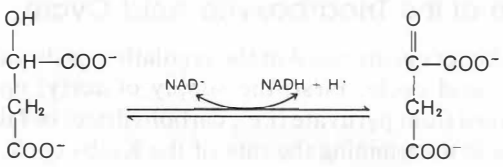
Fumarate L-Malate

tetramer with a mol wt of 200,000 and is stereospecific for the trans form of the substrate (e.g., the cis form, maleate, is not a substrate), and the product of the fumarase reaction is only L-malate. The fumarase reaction is freely reversible under physiological conditions.



Maleate

The final reaction in the Krebs cycle is the malate dehydrogenase reaction in which the final (of three) reducing equivalent as NADH + H^+ is removed from the cycle intermediates.



L-Malate Oxalacetate

The equilibrium of the malate dehydrogenase reaction lies far toward L-malate formation, for example, $\Delta G^\circ = +7.0$ kcal/mol. Thus the malate dehydrogenase reaction is an endothermic reaction when considered in the forward direction of the Krebs cycle. However, the citrate synthase reaction and other reactions of the cycle pull malate dehydrogenase toward oxalacetate formation by removing oxalacetate. Additionally, NADH produced in the various cycle NAD-linked dehydrogenases is rapidly oxidized to NAD^+ in the mitochondrial respiratory chain.

Energy Yield

In summary the tricarboxylic acid cycle (Figure 6.14) serves as a terminal oxidative pathway for most metabolic fuels. Two-carbon moieties as acetyl CoA are taken into the cycle and are oxidized completely to CO_2 and H_2O. During this process 4 reducing equivalents (3 as NADH + H^+ and 1 as $FADH_2$) are produced, which are used subsequently for energy generation. As is discussed later in this chapter; oxidation of each NADH + H^+ results in the formation of 3 ATP molecules in the mitochondrial respiratory chain oxidative phosphorylation sequence, while the oxidation of the $FADH_2$ formed in the succinate dehydrogenase reaction yields 2 ATPs. Also, a high energy bond is formed in the succinyl CoA synthetase reaction. Hence, the net yield of ATP or its equivalent (i.e., GTP) for the complete oxidation of an acetyl group in the Krebs cycle is 12.

During the complete oxidation of glucose to CO_2 and H_2O there is a net formation of (1) 2 ATPs per glucose in the conversion of glucose to two molecules of pyruvate; (2) 6 ATPs per glucose as a result of the translocation and subsequent oxidation in the mitochondrial compartment of 2NADH + H^+ formed in the glyceraldehyde 3-phosphate dehydrogenase reaction of the glycolytic pathway; and (3) 30 ATPs per glucose from the oxidation of the 2 pyruvate molecules in the pyruvate dehydrogenase reaction and subsequent conversion of 2 acetyl CoAs to CO_2 and H_2O in the tricarboxylic acid cycle. Hence, the net ATP yield during the complete oxidation of glucose to $6CO_2$ plus $6H_2O$ is 38 ATPs.

Regulation of the Tricarboxylic Acid Cycle

A variety of factors is involved in the regulation of the activity of the tricarboxylic acid cycle. First, the supply of acetyl units, whether they are derived from pyruvate (i.e., carbohydrate) or fatty acids, is a crucial factor in determining the rate of the Krebs cycle. Regulatory influences on the pyruvate dehydrogenase complex (discussed previously in this chapter) have an important effect on the activity of the cycle. Likewise any control exerted on the processes of transport and β-oxidation of fatty acids would be an effective determinant of the Krebs cycle activity.

Second, because the primary dehydrogenase reactions of the Krebs cycle are dependent upon a continuous supply of both NAD^+ and FAD, their activities are very stringently controlled by the mitochondrial respiratory chain, which is responsible for oxidizing the NADH and $FADH_2$ produced as a result of substrate oxidation in the cycle. Because the activity of the respiratory chain is ob-

ligatorily coupled to the generation of ATP in the oxidative phosphorylation sequence of reactions, the activity of the Krebs cycle is very much dependent upon a "respiratory control," which is strongly affected by the availability of ADP + P_i and oxygen. Hence any inhibitory agent or metabolic condition which might interrupt the supply of oxygen, the continuous supply of ADP or the source of reducing equivalents (e.g., substrate for the cycle) would shut down cycle activity. This type of control of the cycle is generally referred to as the "coarse control" of the cycle. There are, of course, a variety of postulated effector-mediated regulatory interactions between various intermediates or nucleotides and the individual enzymes of the cycle, which may serve to exert a fine control on the activity of the cycle. Some illustrations of these interactions are shown in Figure 6.18 and also have been noted during the discussions of individual enzymes of the Krebs cycle. It must be stressed that the physiological relevance of many of these types of individual regulatory interactions has not been firmly established in intact metabolic systems.

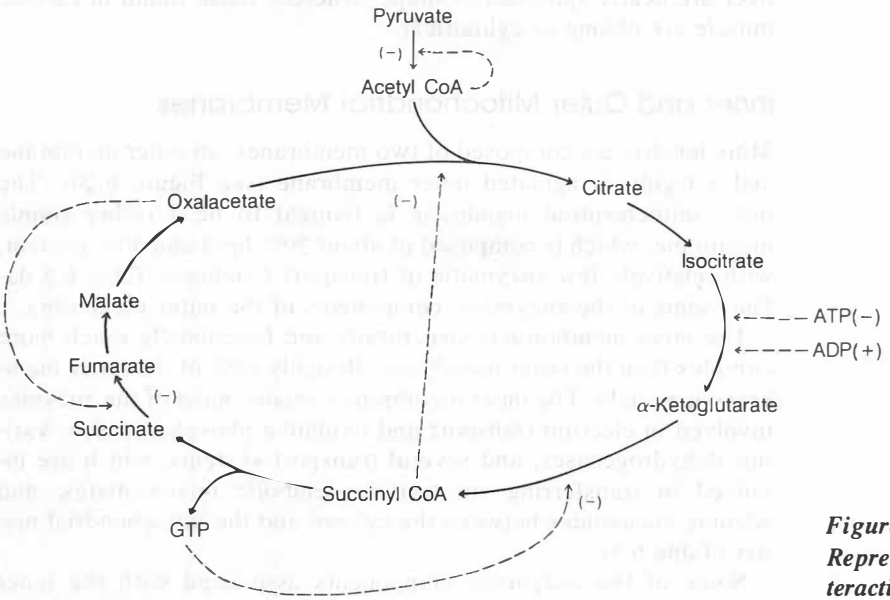


Figure 6.18
Representative examples of the regulatory interactions in the tricarboxylic acid cycle.

6.5 STRUCTURE AND COMPARTMENTATION OF THE MITOCHONDRIAL MEMBRANES

Because the metabolic pathways for the oxidation of pyruvate, the end product of glycolysis, and fatty acids are located in mitochondria, a major portion of the energy-generating capacity of most cells resides in the mitochondrial compartment of the cell. The number of mitochondria in various tissues reflects the physiological function of the tissue and determines its capacity to perform aerobic metabolic functions. For example, the erythrocyte has no mitochondria and hence does not possess the capacity to generate energy, using oxygen as a terminal electron acceptor. On the other hand, cardiac tissue is a highly aerobic tissue, and it has been estimated that about half of the cytoplasmic volume of cardiac cells is composed of mitochondria. The liver is another tissue that is highly dependent upon aerobic metabolic processes for its various functions, and hence it has been estimated that mammalian hepatocytes contain between 800 and 2,000 mitochondria per cell. Mitochondria exist in a variety of different shapes, depending upon the cell type from which they are derived. As can be seen in Figure 6.19 mitochondria from liver are nearly spherical in shape, whereas those found in cardiac muscle are oblong or cylindrical.

Inner and Outer Mitochondrial Membranes

Mitochondria are composed of two membranes, an outer membrane and a highly invaginated inner membrane (see Figure 6.20). The outer mitochondrial membrane is thought to be a rather simple membrane, which is composed of about 50% lipid and 50% protein, with relatively few enzymatic or transport functions. Table 6.5 defines some of the enzymatic components of the outer membrane.

The inner membrane is structurally and functionally much more complex than the outer membrane. Roughly 80% of the inner membrane is protein. The inner membrane contains most of the enzymes involved in electron transport and oxidative phosphorylation, various dehydrogenases, and several transport systems, which are involved in transferring substrates, metabolic intermediates, and adenine nucleotides between the cytosol and the mitochondrial matrix (Table 6.5).

Some of the enzymatic components associated with the inner mitochondrial membrane are only loosely associated with the mem-

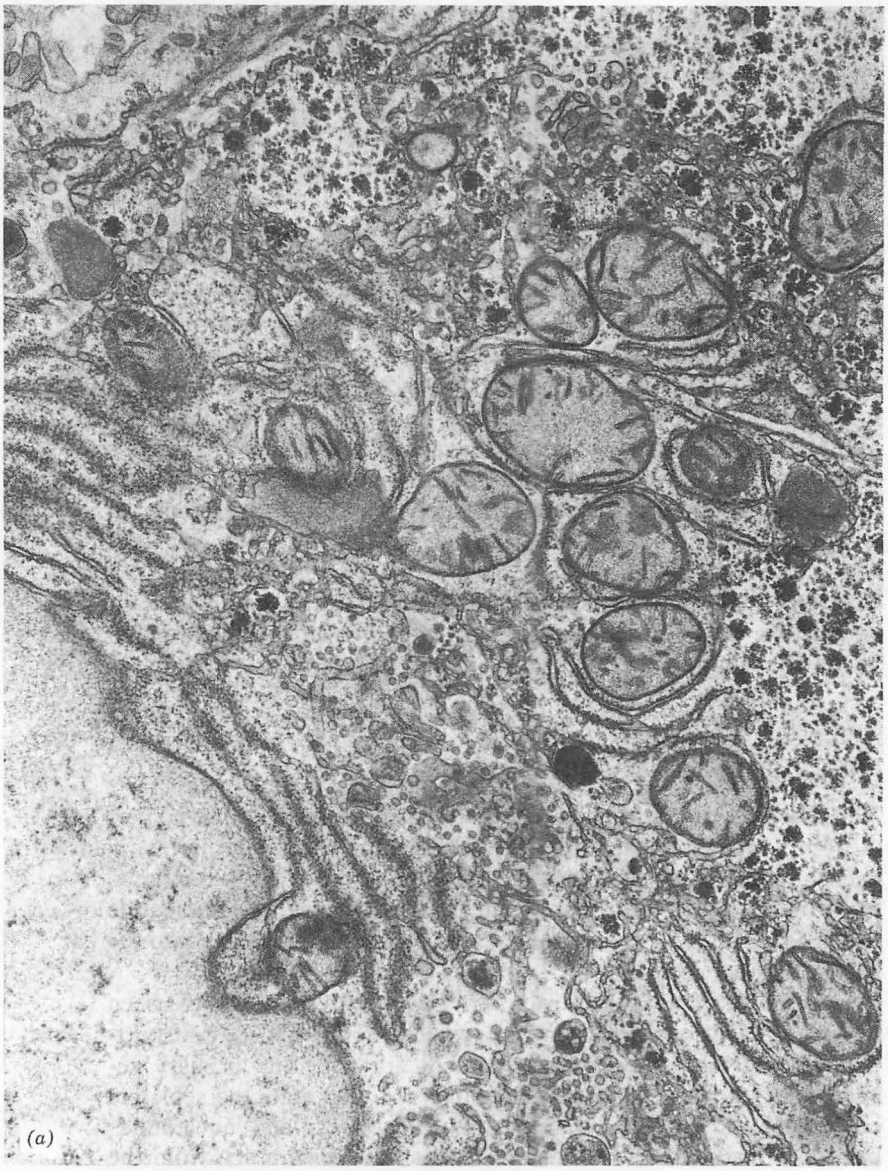
brane, but other enzymatic components are either tightly bound to the membrane or are actual structural elements of the membrane. Hence there is a wide variability in the extent to which physical (ultrasonic irradiation or freezing and thawing), chemical (organic solvent or detergent treatment), or enzymatic (protease or lipase) treatments remove, release, or inactivate the enzymes associated with the inner membrane.

Experimental procedures have been developed that allow the separation of the inner and outer mitochondrial membranes. As indicated in Figure 6.21 the outer membrane may be stripped off and isolated, using digitonin (a detergent), osmotic shock, or ultransonic irradiation followed by density-gradient ultracentrifugation. The resulting inner membrane plus matrix fraction is referred to as a mitoplast. The mitochondrial matrix may be released from the mitoplast, following treatment with a nonionic detergent or vigorous sonication. Once the various subcompartments of the mitochondrion have been separated, analyses may be performed to determine the location of the various characteristic marker enzymes, some of which are listed in Table 6.5. Enzymatic markers have been utilized effectively to detect the presence of mitochondria or even a particular portion of the mitochondrion in membrane preparations of diverse derivation.

Mitochondrial Transport Systems

Whereas the outer mitochondrial membrane presents little or no permeability barrier to substrate or nucleotide molecules of interest in energy metabolism, the inner membrane has very tightly restricted limitations on the types of substrates, intermediates, and nucleotide species that may be transported across the inner membrane into the matrix compartment.

Figure 6.22 depicts the various transport systems that have been described in various mitochondrial systems. Some of these transporters are well characterized, but others are not. The primary responsibility of these transport functions is to facilitate the selective movement of various substrates, intermediates, and nucleotides back and forth across the inner mitochondrial membrane from the cytosol to the mitochondrial matrix. By virtue of having these transporters various substrates and other molecules can be accumulated in the mitochondrial matrix, since the transporters possess the capacity to facilitate the movement of the substrate against a concentration gradient. The importance of the mitochondrial transporter systems derives from their direct involvement in a variety of mitochondrial metabolic processes.



(a)

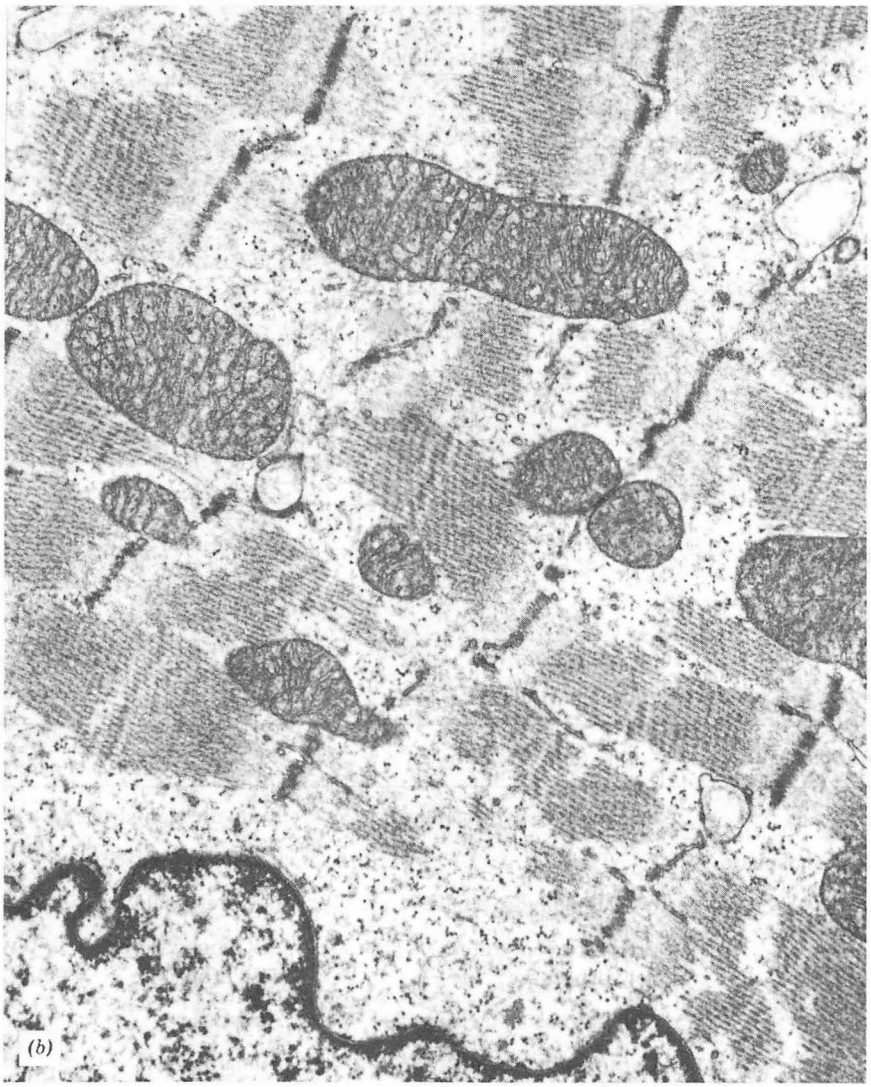


Figure 6.19
Electron micrographs of mitochondria in (a) hepatocytes from rat liver and (b) muscle fibers from rabbit heart.

Magnification 39,600X, courtesy of Dr. W. B. Winborn, Department of Anatomy, The University of Texas Health Science Center at San Antonio, and the Electron Microscopy Laboratory, Department of Pathology, The University of Texas Health Science Center at San Antonio.

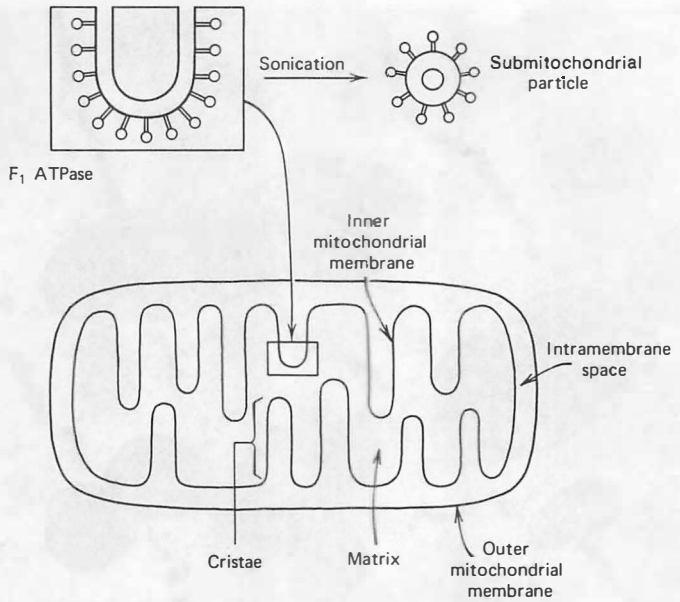


Figure 6.20
Diagram of the various submitochondrial compartments.

Table 6.5 Enzymatic Composition of the Various Mitochondrial Subcompartments

Outer Membrane	*Intramembrane Space*	*Inner Membrane*	*Matrix*
Monoamine oxidase	Adenylate kinase	Succinate dehydrogenase	Pyruvate dehydrogenase
Kynurenine hydroxylase	Nucleoside diphosphate kinase	F_1 ATPase	Citrate synthase
Nucleoside diphosphate kinase		NADH dehydrogenase	Isocitrate dehydrogenase
Phospholipase A		β-Hydroxybutyrate dehydrogenase	α-Ketoglutarate dehydrogenase
Fatty acyl CoA synthetases		Cytochrome b, c_1, c, a, a_3	Aconitase
NADH: cytochrome c reductase (rotenone-insensitive)		Carnitine: acyl CoA transferase	Fumarase
Choline phosphotransferase		Adenine nucleotide translocase	Succinyl CoA synthetase
		Mono-, di-, and tricarboxylate translocase	Malate dehydrogenase
		Glutamate-aspartate translocase	Fatty acid oxidation system
			Glutamate dehydrogenase
			Glutamate-oxalacetate transaminase
			Ornithine transcarbamylase

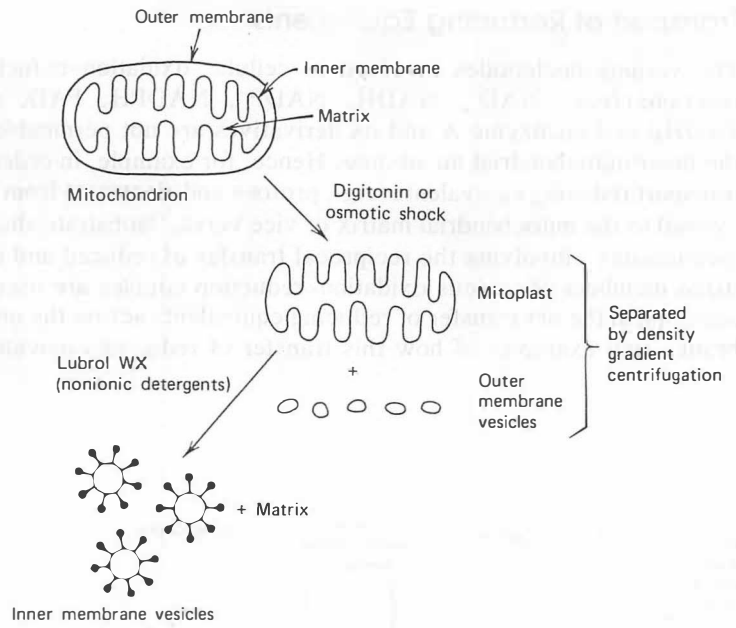


Figure 6.21
Separation of mitochondrial membranes.

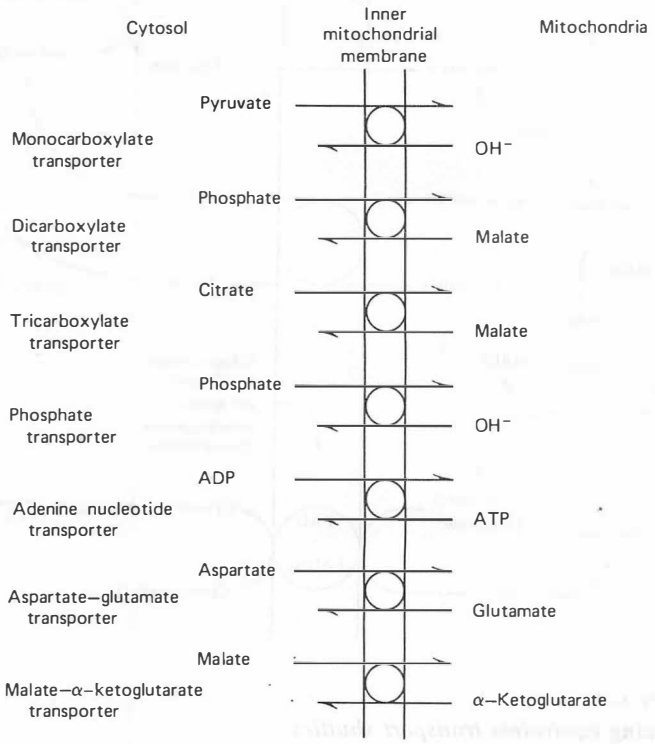


Figure 6.22
Mitochondrial metabolite transporters.

Transport of Reducing Equivalents

The various nucleotides involved in cellular oxidation–reduction reactions (e.g., NAD^+, NADH, $NADP^+$, NADPH, FAD, and $FADH_2$) and coenzyme A and its derivatives are not permeable to the inner mitochondrial membrane. Hence, for example, in order to transport reducing equivalents (e.g., protons and electrons) from the cytosol to the mitochondrial matrix or vice versa, "substrate shuttle mechanisms" involving the reciprocal transfer of reduced and oxidized members of various oxidation–reduction couples are used to accomplish the net transfer of reducing equivalents across the membrane. Two examples of how this transfer of reducing equivalents

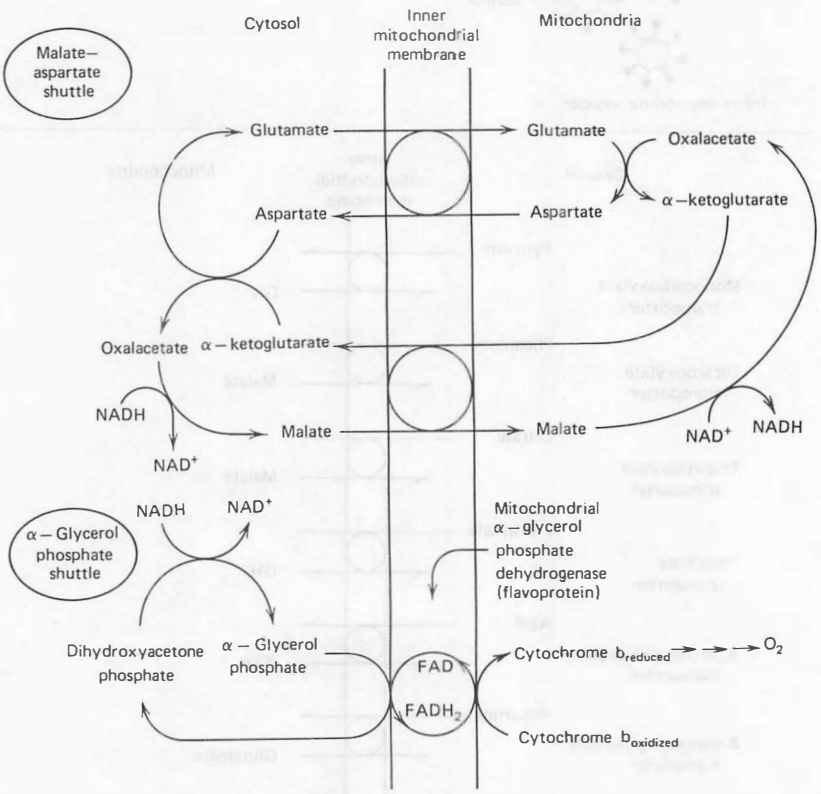


Figure 6.23
Reducing equivalent transport shuttles.

from the cytosol to the mitochondria occurs are shown in Figure 6.23. The malate–aspartate shuttle and the α-glycerol phosphate shuttle are employed in various tissues to translocate reducing equivalents from the cytosol, where they are generated, to the mitochondrial compartment, where they are oxidized to yield energy. The operation of such substrate shuttles requires that the appropriate enzymes are localized on the correct side of the membrane and that appropriate transporters or translocases are present on/in the membrane to shuttle the various intermediates. In this regard the operation of the malate-aspartate shuttle depends on the fact that NADH, NAD^+, and oxalacetate are not permeable to the inner mitochondrial membrane, on the distribution of malate dehydrogenase and the aspartate aminotransferase on both sides of the inner mitochondrial membrane and on the existence of membrane transporters, which allow the exchange of intramitochondrial aspartate for cytosolic glutamate and cytosolic malate for intramitochondrial α-ketoglutarate.

Transport of Acetyl Units

Another example of how an impermeable substance such as acetyl CoA can transfer the 2-carbon fragment (the acetyl group) from the mitochondrial compartment to the cytosol, where acetyl moieties are required for fatty acid or sterol biosynthesis, is illustrated in Figure 6.24.

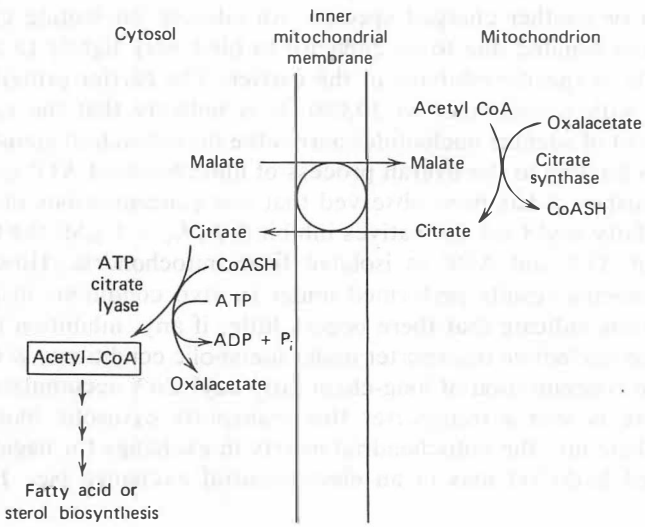


Figure 6.24
The export of intramitochondrially generated citrate to the cytosol to serve as a source of acetyl CoA for biosynthesis of fatty acids or sterols.

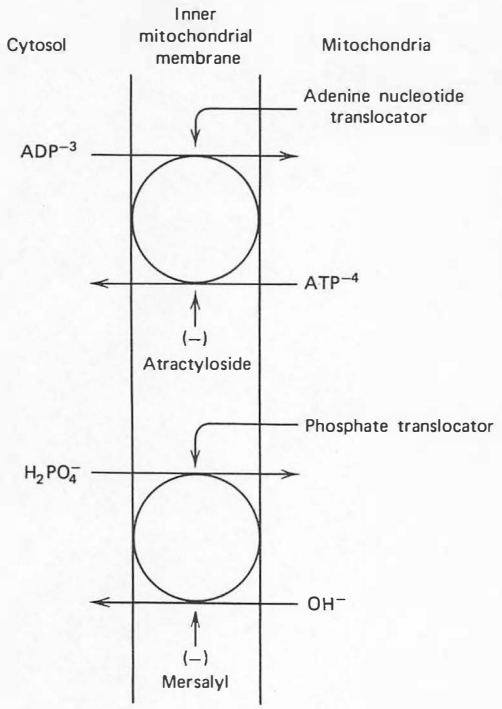


Figure 6.25
The adenine nucleotide and phosphate translocators.

Intramitochondrial acetyl CoA is converted to citrate in the citrate synthase reaction of the Krebs cycle. Subsequently the citrate is exported to the cytosol on the tricarboxylate transporter in exchange for a dicarboxylic acid such as malate. Cytosolic citrate may be cleaved to acetyl CoA and oxalacetate at the expense of an ATP molecule in the ATP citrate lyase reaction, which is discussed in Chapter 9. Substrate shuttle mechanisms also are involved in the movement of appropriate substrates and intermediates in both directions across the inner mitochondrial membrane in the liver during periods of active gluconeogenesis and ureogenesis (see Chapters 7 and 11).

Adenine Nucleotide Transport

Adenine nucleotides are transported across the inner mitochondrial membrane by the very specific adenine nucleotide translocator. Nucleotide species such as the guanine, uridine, or cytosine nucleotides are neither exchanged across the inner membrane on the adenine nucleotide specific translocator nor transported by a comparable carrier specific for nonadenine nucleotides. As indicated in Figure 6.25, cytosolic ADP, which is formed during energy-consuming reactions, is exchanged for mitochondrial ATP, which is generated in the process of oxidative phosphorylation. At pH 7 ADP has three negative charges and ATP has four, so that a 1 : 1 exchange of ADP : ATP would cause a charge imbalance across the membrane. Hence the ADP for ATP exchange across the mitochondrial membrane is an electrogenic process, which requires that in the end the charge imbalance must be compensated for by the movement of a proton or another charged species. An adenine nucleotide carrier has been isolated due to its capacity to bind very tightly to atractyloside, a specific inhibitor of the carrier. The carrier protein is a dimer with subunit mol wt 30,000. It is unlikely that the rate of transport of adenine nucleotides across the mitochondrial membrane ever is limiting to the overall process of mitochondrial ATP synthesis. Further, it has been observed that low concentrations of long-chain fatty acyl CoA derivatives inhibit (i.e., $K_i = 1\ \mu M$) the transport of ATP and ADP in isolated liver mitochondria. However, experimental results performed under in vivo conditions in intact liver cells indicate that there occurs little, if any, inhibition of the adenine nucleotide transporter under metabolic conditions, in which a large concentration of long-chain fatty acyl CoA accumulates.

There is also a transporter that transports cytosolic inorganic phosphate into the mitochondrial matrix in exchange for negatively charged hydroxyl ions in an electroneutral exchange (see Figure

6.24). This phosphate transport may also be accomplished in a proton-compensated mechanism, for example, phosphate and protons are transported in a 1 : 1 ratio. Phosphate transport is strongly inhibited by the compound mersalyl and various mercurial reagents.

Mitochondrial Calcium Transport

Finally, mitochondria from most tissues possess a transport system capable of translocating calcium across the mitochondrial inner membrane. It is difficult to overestimate the importance of the distribution of cellular calcium pools in different cell functions, such as muscle contraction, neural transmission, and hormone action and secretion. Calcium exists in distinct pools in the cell. The cytosol, mitochondria, endoplasmic reticulum, nuclei, and the Golgi membranes have their component pools of intracellular calcium. Some of the intracellular calcium is bound to nucleotides, metabolites, or membrane ligands, while a portion of the intracellular calcium is free in solution. A gradient of calcium exists from outside a cell to inside a cell. Estimates of intracellular (e.g., cytosolic) calcium range from 10^{-6} M in the liver to 10^{-7}–10^{-8} M in the heart and skeletal muscle, whereas extracellular calcium likely is at least two orders of magnitude greater than this. Total intramitochondrial calcium has been estimated to be $\sim 10^{-4}$ M. While the free calcium concentration in the mitochondrion may be in the range of 10^{-5}–10^{-6} M. Hence processes involved in the alternate sequestering and release of an intracellular store of calcium can greatly influence intracellular calcium pools and various cell functions. Mitochondria have been known to accumulate rather large quantities of calcium at the expense of ATP hydrolysis, respiration, or an electrochemical gradient. Mitochondrial calcium transport is inhibited by low concentrations of lanthanides (trivalent metal cations) and by a compound called ruthenium red. Magnesium can compete with calcium for the carrier in certain types of mitochondria. The current view is that there is a specific carrier in the inner mitochondrial membrane, which likely is a glycoprotein. The mitochondrial calcium carrier exhibits saturation kinetics, has a high affinity for calcium, and is highly specific for calcium. Permeant counterions such as phosphate or acetate stimulate calcium transport and allow the metal cation to be retained by the mitochondria. Most interesting is the finding that certain hormones may affect intracellular calcium distribution (e.g., in the mitochondria) as part of the mechanism for the mediation of the hormone response. Various cytosolic protein kinases such as those involved in glycogen metabolism are calcium-sensitive.

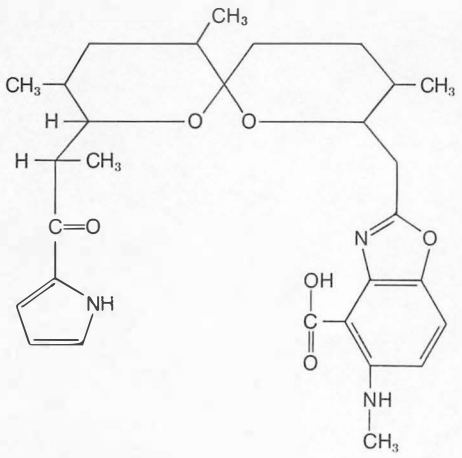


Figure 6.26
Structure of A23187.

Ionophores

The translocation of divalent metal cations such as calcium or magnesium across cellular membranes can be facilitated by a group of compounds called ionophores. The divalent metal cation-specific ionophores A23187 (Figure 6.26) and X537A may be used to move calcium across membranes down concentration gradients. Alkali metal cations such as potassium are complexed by a group of depsipeptide ionophorous compounds. Valinomycin is one such compound, which binds potassium and moves it down concentration gradients across the lipophilic portions of membranes. Most ionophores accomplish their transport function because they are apolar compounds that can selectively bind to metal cations and facilitate the movement of the polar metal cation across the lipophilic portions of natural membranes.

In summary, the inner mitochondrial membrane possesses a variety of transport systems, that are involved in the movement of nucleotides, substrates, metabolites, and metal cations into and out of the mitochondrial compartment. An understanding of these transport functions is essential in order to understand complex cellular metabolic pathways and their regulation. (See Clin. Corr. 6.2.)

CLIN. CORR. 6.2 MITOCHONDRIAL MYOPATHIES

Muscle myopathies that involve defects in various metabolic functions of muscle have been described. Clinically, patients with muscle myopathies complain of weakness and cramping of the affected muscles; infants have difficulty feeding and crawling; severe fatigue results from minimal exertion; and there is usually evidence of muscle wasting. On the basis of electron microscopic examination and enzymatic characterization of muscle biopsy material, many myopathies have been found that have a primary lesion in mitochondrial function.

6.6 ELECTRON TRANSFER

During the enzymatic reactions involved in glycolysis, fatty acid oxidation, and the tricarboxylic acid cycle, reducing equivalents are derived from the sequential breakdown of the initial metabolic fuel. In the case of glycolysis, NADH is produced in the glyceraldehyde 3-phosphate dehydrogenase reaction, and this reducing equivalent must be either reoxidized in the cytosol (e.g., by lactate dehydrogenase) or transported to the mitochondrial matrix via one of the substrate shuttle mechanisms in order to realize the maximum energy yield from the oxidation of glucose. In the case of fatty acid oxidation and in the tricarboxylic acid cycle, reducing equivalents as both NADH and $FADH_2$ are produced in the mitochondrial matrix. In order to transduce this reducing power into utilizable energy, mitochondria have a system of electron carriers in or associated with the inner mitochondrial membrane, which convert reducing equivalents in the presence of oxygen into utilizable energy by synthesizing ATP. This process is called electron transport, and, as will be seen later, NADH and $FADH_2$ oxidation in this process results in the production of 3 and 2 mol ATP/mol reducing equivalent transferred to oxygen, respectively.

Oxidation–Reduction Reactions

Prior to the presentation of a description of the many components and the mechanism of the electron transport sequence, it is important to discuss some basic information concerning oxidation–reduction reactions. The mitochondrial electron transport system is little more than a sequence of linked oxidation–reduction reactions, for example,

$$AH_2 + B \rightleftharpoons A + BH_2$$

or

$$\text{Electron donor} \rightleftharpoons \text{electron} + \text{electron acceptor}$$

Oxidation–reduction reactions occur when there is a transfer of electrons from a suitable electron donor (the reductant) to a suitable electron acceptor (the oxidant). In some oxidation–reduction reactions only electrons are transferred from the reductant to the oxidant (i.e., electron transfer between cytochromes),

$$\text{Cytochrome c } (Fe^{2+}) + \text{cytochrome a } (Fe^{3+}) \rightleftharpoons \text{cytochrome c } (Fe^{3+}) + \text{cytochrome a } (Fe^{2+})$$

whereas in other types of reactions, both electrons and protons (hydrogen atoms) are transferred (i.e., electron transfer between NADH and FAD).

$$NADH + H^+ + FAD \rightleftharpoons NAD^+ + FADH_2$$

The oxidized and the reduced forms of the compounds or groups operating in oxidation–reduction-type reactions are referred to as redox couples or pairs. The facility with which a given electron donor (reductant) gives up its electrons to an electron acceptor (oxidant) is expressed quantitatively as the oxidation–reduction potential of the system. An oxidation–reduction potential is measured as an electromotive force in volts of a half-cell made up of both members of an oxidation–reduction couple when compared to a standard reference half-cell (usually the hydrogen electrode reaction) (see Figure 6.27). The potential of the standard hydrogen reference electrode is set by convention at 0.0 V at pH 0.0. However, when this standard potential is corrected for pH 7.0 the reference electrode potential becomes −0.42 V. The oxidation–reduction potentials for a variety of important biochemical reactions have been determined and are tabulated in Table 6.6.

Deficiencies in mitochondrial transport functions (i.e., carnitine: palmitoyl CoA transferase) and in components of the mitochondrial electron transport chain (NADH dehydrogenase, cytochrome b, cytochrome a,a_3, or the mitochondrial ATPase) have been described. In many mitochondrial myopathies large paracrystalline inclusions have been found within the mitochondrial space. (See Clin. Corr. 6.2 Figure.) It is not known whether this crystalline material is inorganic or organic in composition. In certain mitochondrial myopathies electron transport is only loosely coupled to ATP production; in other cases these processes exhibit normal tight coupling.

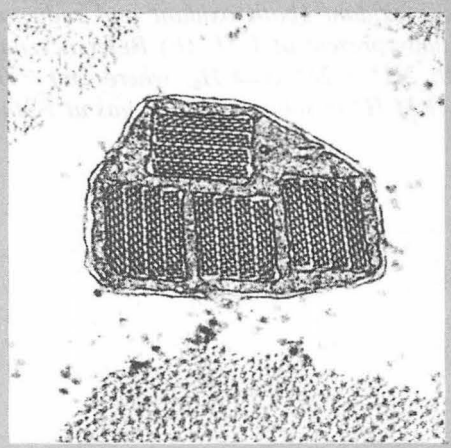


Figure CC6.2
Example of paracrystalline inclusions in mitochondria from muscles of ocular myopathic patients.
Magnification 36,000X, courtesy of Dr. D. N. Landon, Institute of Neurology, University of London.

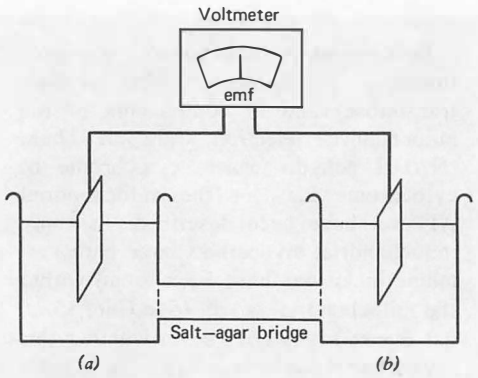


Figure 6.27
A system for determining oxidation–reduction potentials in half-cell reactions.
(a) Sample half-cell; Red:H + Ox ⇌ Ox:H + Red, where Red = reductant and Ox = oxidant (both oxidant and reductant initially present at 1 M. (b) Reference half-cell; $2H^+ + 2e^- \rightleftharpoons H_2$, where emf = 0.0 V (1 M H^+ in solution and H_2 gas at 1 atm).

Table 6.6 Standard Oxidation–Reduction Potentials for Various Biochemical Reactions

Oxidation–Reduction System	*Standard Oxidation–Reduction Potential E_0' (V)*
Acetate + 2H + $2e^-$ ⇌ acetaldehyde	−0.60
$2H^+ + 2e^-$ ⇌ H_2	−0.42
Acetoacetate + $2H^+ + 2e^-$ ⇌ β-hydroxybutyrate	−0.35
$NAD^+ + 2H^+ + 2e^-$ ⇌ NADH + H^+	−0.32
Acetaldehyde + $2H^+ + 2e^-$ ⇌ ethanol	−0.20
Pyruvate + $2H^+ + 2e^-$ ⇌ lactate	−0.19
Oxalacetate + $2H^+ + 2e^-$ ⇌ malate	−0.17
Coenzyme Q_{ox} + $2H^+ + 2e^-$ ⇌ coenzyme Q_{red}	+0.10
Cytochrome b (Fe^{3+}) + e^- ⇌ cytochrome b (Fe^{2+})	+0.12
Cytochrome c (Fe^{3+}) + e^- ⇌ cytochrome c (Fe^{2+})	+0.22
Cytochrome a (Fe^{3+}) + e^- ⇌ cytochrome a (Fe^{2+})	+0.29
$\frac{1}{2}O_2 + 2H^+ + 2e^-$ ⇌ H_2O	+0.82

An important concept is indicated in this listing of oxidation–reduction potentials. The reductant of an oxidation–reduction pair with large negative oxidation–reduction potential will give up its electrons more readily than pairs with smaller negative or positive redox potentials. On the other hand, a strong oxidant (e.g., characterized by a large positive potential) has a very high affinity for electrons.

The Nernst equation characterizes the relationship between the standard oxidation–reduction potential of a particular redox pair (E_0'), the observed potential (E), and the ratio of the concentrations of the oxidant and reductant in the system:

$$E = E_0' + \frac{2.3\,RT}{nF} \log \frac{[\text{oxidant}]}{[\text{reductant}]}.$$

E is the observed potential with all concentrations at 1 M.

E_0' is the standard potential at pH 7.0.

R is the gas constant of 8.3 J deg^{-1} mol^{-1}.

T is the absolute temperature in Kelvins.

n is the number of electrons being transferred.

F is the Faraday of 96,500 J V^{-1}.

When the observed potential is equal to the standard potential, a potential is defined which is referred to as the midpoint potential. At the midpoint potential the concentration of the oxidant is equal to the reductant. Knowing the standard oxidation–reduction potentials of a diverse variety of biochemical reactions allows one to predict the direction of electron flow or transfer when more than one redox pair is linked together by the appropriate enzyme, which causes a reaction to occur. For example, as shown in Table 6.6 the $NAD^+/NADH$ pair has a standard potential of -0.32 V, and the pyruvate/lactate pair possesses a potential of -0.19. This means that electrons will flow from the $NAD^+/NADH$ system to the pyruvate/lactate system as long as the enzymatic component (lactate dehydrogenase) is present, for example,

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{lactate} + \text{NAD}^+$$

Hence in the mitochondrial electron transfer system electrons or reducing equivalents are being produced in the primary NAD^+- and FAD-linked dehydrogenase reactions, which have standard electrode potentials at or close to the $NAD^+/NADH$ oxidation–reduction pair and are being passed through the electron transfer chain, which has as its terminal acceptor the oxygen/water oxidation–reduction couple.

Free Energy Changes in Redox Reactions

Oxidation–reduction potential differences between two redox pairs are similar to free energy changes in a chemical reaction, in that both quantities depend on the concentration of the reactants and products of the reaction. Because of this similarity the following relationship can be expressed:

$$\Delta G^{\circ\prime} = -nF\ \Delta E_0'$$

Using this expression the free energy change for electron transfer reactions can be readily calculated if the potential difference between two oxidation–reduction pairs is known. Hence for the mitochondrial electron transfer process in which electrons are transferred between the $NAD^+/NADH$ couple ($E_0' = -0.32$ V) and the $\frac{1}{2}O_2/H_2O$ couple ($E_o' = +0.82$ V) the free energy change for this process can be calculated:

$$\Delta G^\circ = nF\ \Delta E_0' = -2 \times 23.062 \times 1.14 \text{ V}$$

$$\Delta G^\circ = -52.6 \text{ kcal/mol}$$

where 23.062 is the Faraday in kcal V^{-1}. Thus the free energy available from the potential span between NADH and oxygen in the electron transfer chain is capable of generating more than enough energy to synthesize three ATPs per two reducing equivalents or two electrons transported to oxygen. Additionally, because of the negative sign of the free energy available in the process of mitochondrial electron transfer, this process is exergonic and will proceed spontaneously provided that the necessary enzymatic components are present.

Major Components of the Mitochondrial Electron Transport Chain

Prior to cataloging the mechanistic details of the mitochondrial electron transport chain it is necessary to describe the various components that participate in the transfer of electrons in this system. The major enzymes or proteins functioning as electron transfer components involved in the mitochondrial electron transfer system are as follows:

1. NAD^+-linked dehydrogenases
2. Flavin-linked dehydrogenases
3. Iron-sulfur proteins
4. Cytochromes

NAD^+-Linked Dehydrogenases

The initial stage in the mitochondrial electron transport sequence consists of the generation of reducing equivalents in the tricarboxylic acid cycle, the fatty acid β-oxidation sequence, and various other dehydrogenase reactions. The NAD-linked dehydrogenase reactions of these pathways reduce NAD^+ to NADH while converting the reduced member of an oxidation–reduction couple to the oxidized form, for example, for the isocitrate dehydrogenase reaction

$$\text{Isocitrate} + NAD^+ \rightleftharpoons \alpha\text{-ketoglutarate} + CO_2 + NADH + H^+$$

There are two forms of pyridine nucleotides involved in various metabolic reactions, for example, NAD^+ and $NADP^+$ (Figure 6.28). The only difference in these two species is the fact that $NADP^+$ has a phosphate moiety esterified to the 2 position of the ribose in the adenosine portion of the dinucleotide. Each $NAD(P)^+$-linked dehydrogenase catalyzes a stereospecific transfer of the reducing equivalent from the substrate to the nucleotide:

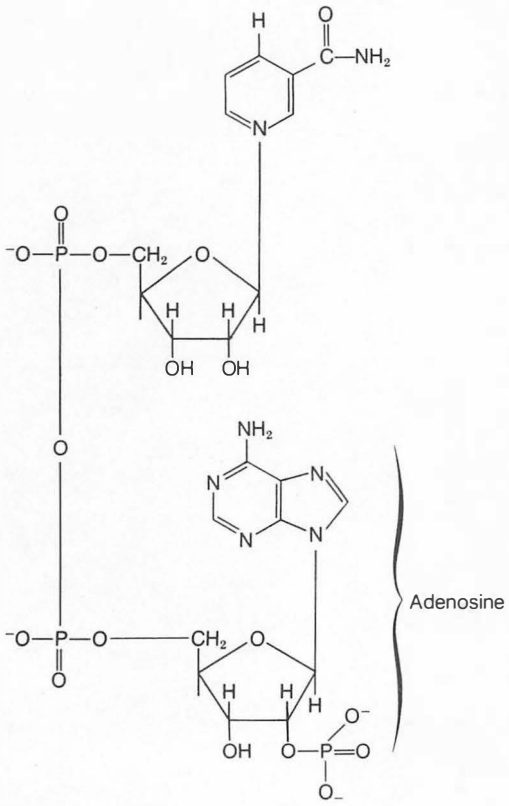
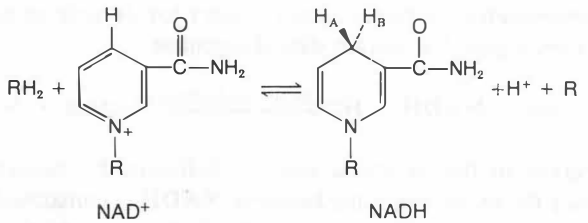


Figure 6.28
Structure of nicotinamide adenine dinucleotide phosphate: $NADP^+$.



$NAD(P)^+$-linked dehydrogenases are either A-specific or B-specific in that the transfer of hydrogen occurs between either the oxidized or reduced metabolite and the A-side (projecting out from the plane of the pyridine ring) or the B-side (below the plane of the ring). Table 6.7 lists several examples of the stereospecificity of $NAD(P)^+$-linked dehydrogenases. Once formed, NAD(P)H is released from the primary dehydrogenase and serves as the substrate for the next step in the mitochondrial electron transport system. NADPH, which is generated in certain dehydrogenase reactions, is not a substrate for the mitochondrial respiratory chain but is usually the form of the reducing equivalents utilized in the reductive biosynthetic reactions of such processes as fatty acid and sterol synthesis. When $NAD(P)^+$ is converted to NAD(P)H, there is a characteristic change in the absorbent and fluorescent properties of these nucleotides, which occurs as a result of the reduction of $NAD(P)^+$. As seen in Figure 6.29, the reduced form of the pyridine nucleotide [NAD(P)H] has an absorbance maximum at 340 nm, not present in the oxidized $NAD(P)^+$ form. Further, when the reduced form of the pyridine nucleotide is excited by light at 340 nm a fluorescence emission maximum is seen at 465 nm. These absorbent and fluorescent properties of the pyridine nucleotides have been employed extensively in developing

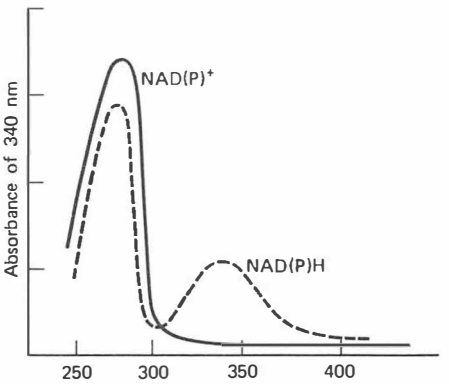


Figure 6.29
Absorbance properties of NAD^+ and NADH.

Table 6.7 The Stereospecificity of $NAD^+(P)$-Linked Dehydrogenases

$NAD^+(P)$-Linked Dehydrogenase	*Specificity*
Alcohol dehydrogenase	A
Malate dehydrogenase	A
Lactate dehydrogenase	A
Isocitrate dehydrogenase ($NADP^+$)	A
Hydroxylacyl CoA dehydrogenase	B
Glyceraldehyde 3-phosphate dehydrogenase	B
Glucose 6-phosphate dehydrogenase ($NADP^+$)	B

spectrophotometric or fluorescence assays for dehydrogenase reactions, for example, for lactate dehydrogenase,

$$\text{Pyruvate} + \text{NADH} + \text{H}^+ \xrightleftharpoons{\text{lactate dehydrogenase}} \text{lactate} + \text{NAD}^+$$

The progress of this reaction may be followed by measuring the absorbency decrease with time because NADH is consumed, using a spectrophotometer set at a wavelength of 340 nm. Additionally, the absorbent and fluorescent properties of the pyridine nucleotides have been utilized to monitor the oxidation–reduction state of a tissue or a preparation of intact mitochondria. With an appropriate spectrophotometer (e.g., a dual wavelength spectrophotometer), which is capable of measuring small absorbency changes in turbid cell or mitochondrial suspensions, the relative changes in the oxidized/reduced pyridine nucleotides may be determined as a function of the metabolic condition of the cell or subcellular suspension (e.g., changes in substrate, oxygen concentration, or drug or hormone additions). This type of spectrophotometric technique and more sophisticated techniques, in which a light guide is used to direct a beam of excitation light to the surface of an intact organ or tissue, and another light guide is employed to pick up the reflected fluorescence emission at a longer wavelength, have been valuable tools in understanding the very complicated relationships that exist between the mitochondrial respiratory chain and the metabolic characteristics of various tissues.

Finally, another effective method for selectively monitoring the oxidation–reduction state of the cytosolic or the mitochondrial compartments in an intact tissue is to measure the oxidized and reduced members of various redox couples in tissue extracts, in the bathing solution of a given tissue or in the effluent perfusate of an isolated, perfused organ. Because the lactate dehydrogenase is exclusively a cytosolic enzyme the pyruvate/lactate ratio in the tissue or organ perfusate should accurately reflect the cytosolic $NAD^+/NADH$ ratio under a variety of metabolic conditions. In a like manner the β-hydroxybutyrate dehydrogenase is exclusively a mitochondrial enzyme, and hence the acetoacetate/β-hydroxybutyrate ratio should reflect the oxidation–reduction state of the mitochondrial $NAD^+/NADH$ system. If the acetoacetate/β-hydroxybutyrate ratio and the equilibrium constant for the β-hydroxybutyrate dehydrogenase are known, the $NAD^+/NADH$ ratio under any condition can be calculated:

$$\text{Acetoacetate} + \text{NADH} + \text{H}^+ \rightleftharpoons \beta\text{-hydroxybutyrate} + \text{NAD}^+$$

$$K_{eq} = \frac{[\beta\text{-hydroxybutyrate}][\text{NAD}^+]}{[\text{acetoacetate}][\text{NADH}][\text{H}^+]}$$

Flavin-Linked Dehydrogenases

The second type of oxidation–reduction reaction essential to a discussion of mitochondrial electron transport employs a flavin (e.g., derived from riboflavin 5′-phosphate) as the electron acceptor in the reaction. These reactions are catalyzed by a group of flavin-linked dehydrogenases. The two flavins commonly utilized in oxidation–reduction reactions are FAD (flavin adenine dinucleotide) and FMN (flavin mononucleotide) (Figure 6.30).

Among the wide variety of flavin-containing enzymes that have been described in nature are four or five that play an essential role in energy metabolism in mammalian mitochondria (Table 6.8). In the preceding discussion of the pyruvate and the α-ketoglutarate dehydrogenase multienzyme complexes the final portion of the reaction catalyzed by this complex involved the flavoprotein enzyme, dihydrolipoyl dehydrogenase, which accepts electrons via a bound FAD moiety from reduced lipoamide groups on the transacylase subunit and transfers these reducing equivalents to NAD^+. Also, in the tricarboxylic acid cycle the succinate dehydrogenase is a flavin-linked enzyme, which oxidizes succinate to fumarate and converts FAD to $FADH_2$.

The first dehydrogenation reaction in the β-oxidation of fatty acids is catalyzed by the acyl CoA dehydrogenase, another flavin-linked enzyme. Finally, the initial oxidation of NADH in the mitochondrial respiratory chain is catalyzed by a FMN-containing enzyme, the NADH dehydrogenase, and the reducing equivalents are then transferred to another flavoprotein called the electron-transferring flavo-

Table 6.8 Various Flavin-Linked Dehydrogenases

Enzyme	*Function*	*Flavin Nucleotide*
Succinate dehydrogenase	Tricarboxylic acid cycle	FAD
Dihydrolipoyl dehydrogenase	Component in pyruvate and α-ketoglutarate dehydrogenases	FAD
NADH dehydrogenase	Electron transport chain	FMN
Electron-transferring flavoprotein	Electron transport chain	FAD
Acyl CoA dehydrogenase	Fatty acid oxidation	FAD
D-Amino acid oxidase	Amino acid oxidation	FAD
Monoamine oxidase	Oxidation of monoamines	FAD

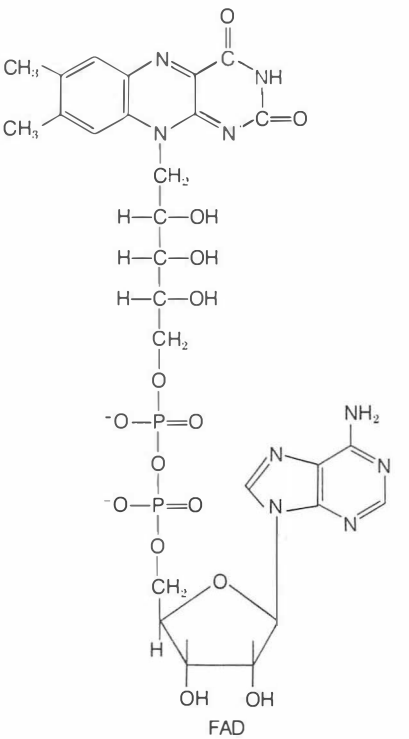


Figure 6.30
Structures of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN).

protein. The flavins FAD and FMN either may be bound very tightly with noncovalent bonding (i.e., with dissociation constants in the range of 10^{-10} M) to their respective enzymes, as is the case with the NADH dehydrogenase, or the flavins may be bound covalently to the enzyme (i.e., to a histidine residue), as is the case with succinate dehydrogenase. Flavoproteins may be classified into two groups: (1) the dehydrogenases in which the reduced flavin is reoxidized by electron carriers other than oxygen (e.g., coenzyme Q, other flavins, or chemicals such as ferricyanide, methylene blue, or phenazine methosulfate), and (2) the oxidases in which the flavin may be reoxidized using as the electron acceptor molecular oxygen, O_2, yielding hydrogen peroxide, H_2O_2, as the product. The H_2O_2 may then be broken down to water and oxygen by the enzyme catalase,

$$2H_2O_2 \xrightarrow{\text{catalase}} 2H_2O + O_2$$

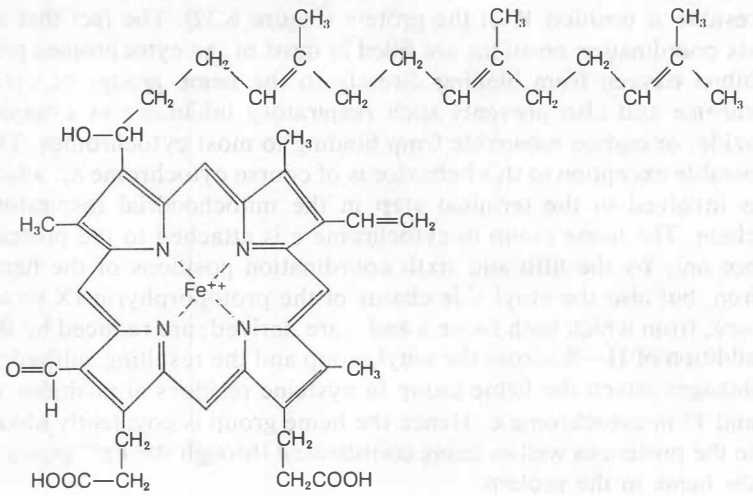
Iron-Sulfur Proteins

A number of flavin-linked enzymes have nonheme iron (i.e., an iron-sulfur center) involved in the catalytic mechanism. In these enzymes the iron is converted from the oxidized (Fe^{3+}) form to reduced (Fe^{2+}) form during the transfer of reducing equivalents on and off the flavin moiety. Both succinate dehydrogenase and NADH dehydrogenase contain iron-sulfur centers. The iron component of the iron-sulfur center is bound in various arrangements to cysteine residues in the protein and to acid-labile sulfur, for example, $Fe_4S_4Cys_4$; $Fe_2S_2Cys_4$; $Fe_1S_0Cys_4$. The iron in the iron-sulfer center undergoes interconversion from the Fe^{2+} (reduced) to the Fe^{3+} (oxidized) state during the participation of these proteins in electron transfer reactions. Iron-sulfur proteins are found in abundance in all species from the simplest microorganism to the mammal. Additionally, certain flavin-linked enzymes have one or two molybdenum atoms associated with their catalytic mechanism. The tightly bound molybdenum undergoes a valence change during the transfer of electrons within the flavin-enzyme system, $Mo^{+6} \longrightarrow Mo^{+5}$.

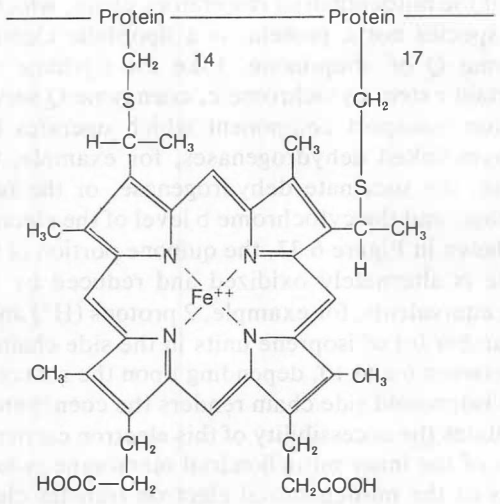
Cytochromes

Finally, organisms that require oxygen (i.e., aerobic organisms) in their energy-generating functions possess various cytochromes that are involved in electron-transferring systems. Cytochromes are a class of proteins characterized by the presence of an iron-containing heme group covalently bound to the protein. Unlike the heme group in hemoglobin or myoglobin in which the heme iron remains in the Fe^{2+} state, the iron in the heme of a cytochrome alternately is oxidized Fe^{3+} or reduced (Fe^{2+}) as the cytochrome transfers electrons toward oxygen in the electron transport chain.

The cytochromes of mammalian mitochondria were designated as a, b, and c on the basis of the α band of the absorption spectrum of the cytochrome. Also, the a, b, and c designation refers to the type of hemin group in the cytochrome (see Figure 6.31). One of the most common electron-transferring cytochromes is cytochrome c. This is



Heme A



Heme C

Figure 6.31
Structure of heme A and heme C.

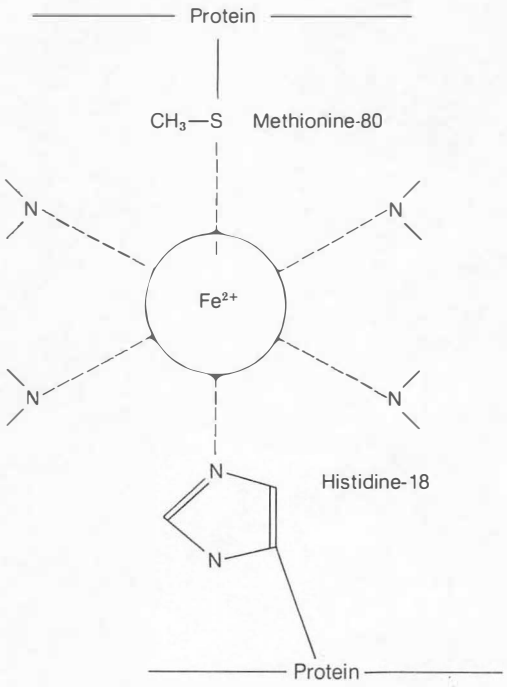


Figure 6.32
The six coordination positions of cytochrome c.

a small protein (104 amino acid residues) with mol wt = 13,000. Amino acid sequences of cytochrome c from a great many species have been performed, and it has been shown that the positions of 20 out of 104 amino acid residues are invariant. It has been demonstrated that the iron of the heme group in cytochrome c is coordinated between the four nitrogens of the tetrapyrrole structure of the porphyrin group, while the fifth and sixth coordination positions are occupied by a methionine residue at position 80 and a histidine residue at position 18 of the protein (Figure 6.32). The fact that all six coordination positions are filled in most of the cytochromes prohibits oxygen from binding directly to the heme groups of cytochrome and also prevents such respiratory inhibitors as cyanide, azide, or carbon monoxide from binding to most cytochromes. The notable exception to this behavior is of course cytochrome a_3, which is involved in the terminal step in the mitochondrial respiratory chain. The heme group in cytochrome c is attached to the protein, not only by the fifth and sixth coordination positions of the heme iron, but also the vinyl side chains of the protoporphyrin IX structure, from which both heme a and c are derived, are reduced by the addition of H—S across the vinyl group and the resulting sulfhydryl linkages attach the heme group to cysteine residues at positions 14 and 17 in cytochrome c. Hence the heme group is covalently linked to the protein as well as being coordinated through the Fe^{2+} group in the heme to the protein.

Coenzyme Q

The final component of the mitochondrial respiratory chain, which is neither a nucleotide species nor a protein, is a lipophilic electron carrier called coenzyme Q or ubiquinone. Like the pyridine nucleotides and to a certain extent cytochrome c, coenzyme Q serves as a "mobile" electron transport component which operates between the various flavin-linked dehydrogenases, for example, the NADH dehydrogenase, the succinate dehydrogenase, or the fatty acyl CoA dehydrogenase, and the cytochrome b level of the electron transport chain. As shown in Figure 6.33, the quinone portion of the coenzyme Q molecule is alternately oxidized and reduced by the addition of 2 reducing equivalents, for example, 2 protons (H^+) and 2 electrons (e^-). The number (n) of isoprene units in the side chain of coenzyme Q varies between 6 and 10, depending upon the source of the coenzyme Q. The isoprenoid side chain renders the coenzyme Q lipid-soluble and facilitates the accessibility of this electron carrier to the lipophilic portions of the inner mitochondrial membrane, where the enzymatic aspects of the mitochondrial electron transfer chain are localized.

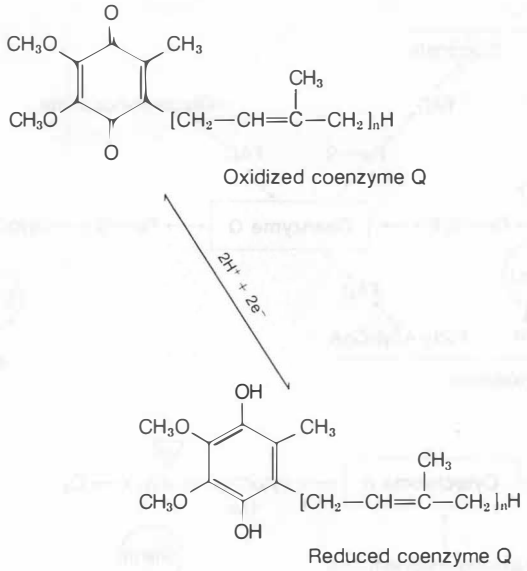


Figure 6.33
The oxidation and reduction of coenzyme Q.

Mitochondrial Electron Transport Chain

The various electron-transferring proteins and other carriers that comprise the mitochondrial electron transfer chain are arranged in a sequential pattern in the inner mitochondrial membrane. Reducing equivalents are extracted from substrates in the tricarboxylic acid cycle, the fatty acid β-oxidation sequence and indirectly from glycolysis and passed sequentially through the electron transport chain to molecular oxygen. The current arrangement of the mitochondrial electron transport carriers is illustrated in Figure 6.34. Electrons or reducing equivalents are fed into the electron transport chain at the level of NADH or coenzyme Q from the primary NAD^+- and FAD-linked dehydrogenase reactions and are transported to molecular oxygen through the cytochrome chain. This electron transport system is set up so that the reduced member of one redox couple is oxidized by the oxidized member of the next component in the system:

$$NADH + H^+ + FMN \rightleftharpoons FMNH_2 + NAD^+$$

or

$$\text{Cytochrome b } (Fe^{2+}) + \text{cytochrome } c_1\ (Fe^{3+}) \rightleftharpoons \text{cytochrome b } (Fe^{3+}) + \text{cytochrome } c_1\ (Fe^{2+})$$

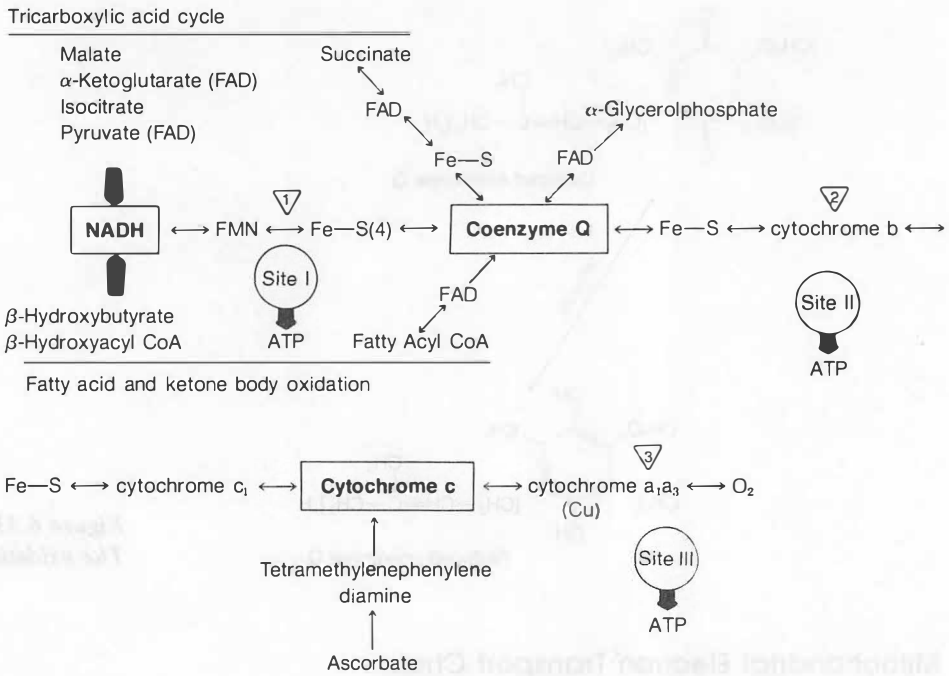


Figure 6.34
Mitochondrial electron transport chain.

The various components of the respiratory chain have characteristic absorption spectra which can be visualized in suspensions of isolated mitochondria or submitochondrial particles using a dual beam spectrophotometer. The different absorption bands are shown in Figure 6.35. One of the light beams of the spectrophotometer was passed through a suspension of liver mitochondria, which was maintained under fully reduced conditions (e.g., substrate plus no oxygen), and the other beam was passed through an identical suspension in the presence of oxygen. Hence the resulting spectra is actually a difference spectrum of the oxidized and reduced states of the mitochondrial respiratory chain.

During the transfer of electrons from the NADH/NAD$^+$ couple ($E_0' = -0.32$) to molecular oxygen ($E_0' = +0.82$) there occurs an oxidation–reduction potential decrease of 1.14 V. As shown in Figure 6.36 this drop in potential occurs in discrete steps as reducing equivalents or electrons are passed between the different segments of the chain. There is at least a 0.3 V decrease in potential between each of the three coupling or phosphorylation sites. A potential drop

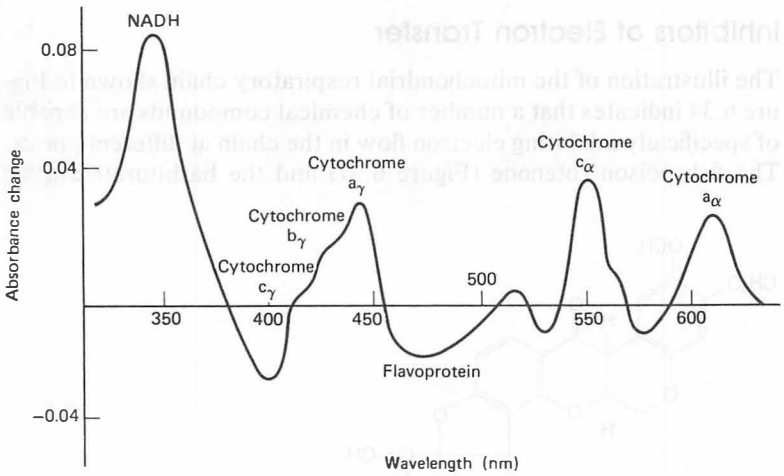


Figure 6.35
Difference spectra of liver mitochondrial suspensions (oxidized/reduced).

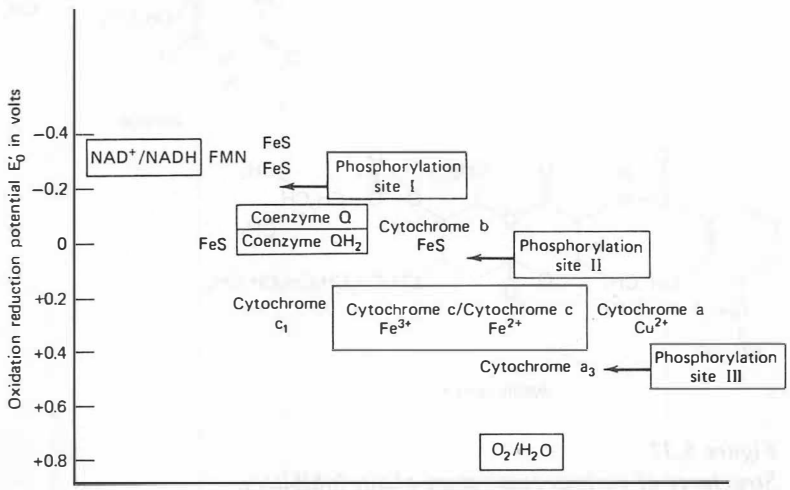


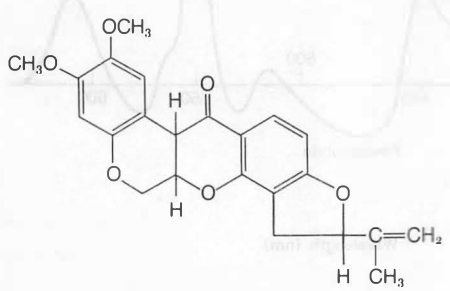
Figure 6.36
Oxidation–reduction potential plot of the mitochondrial electron transport chain carriers.

of 0.3 V is more than sufficient to accommodate the synthesis of a high energy phosphate bond such as occurs in ATP synthesis, for example,

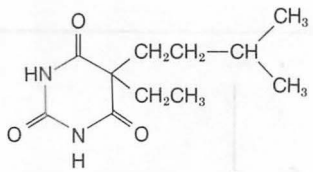
$$\Delta E_0' = 0.3 \text{ V}$$
$$\Delta G^\circ = nF\Delta E_0'$$
$$\Delta G^\circ = 2 \times 23.062 \times 0.3$$
$$\Delta G^\circ = -13.8 \text{ Kcal/mol}$$

Inhibitors of Electron Transfer

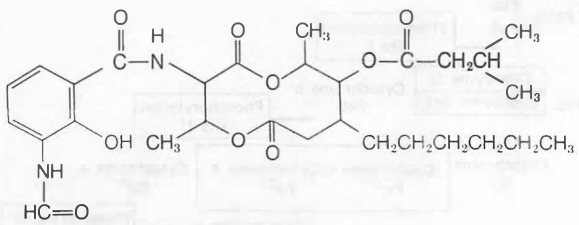
The illustration of the mitochondrial respiratory chain shown in Figure 6.34 indicates that a number of chemical compounds are capable of specifically inhibiting electron flow in the chain at different points. The fish poison rotenone (Figure 6.37) and the barbiturate amytal



Rotenone



Amytal



Antimycin A

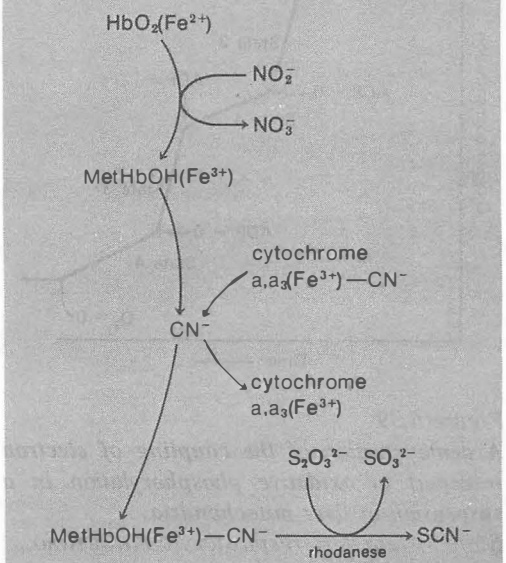
Figure 6.37
Structures of various respiratory chain inhibitors.

CLIN. CORR. 6.3 CYANIDE POISONING

Inhalation of hydrogen cyanide gas or ingestion of potassium cyanide causes a rapid and extensive inhibition of the mitochondrial electron transport chain at the cytochrome oxidase step. Cyanide is one of the most potent and rapidly acting poisons known. Cyanide binds to the Fe^{3+} in the heme of the cytochrome a,a_3 component of the terminal step in electron transport chain and prevents oxygen from reacting with the cytochrome a,a_3. Mitochondrial respiration and energy production cease, and cell death occurs rapidly. Death due to cyanide poisoning occurs from tissue asphyxia, most notably of the central nervous system. If cyanide poisoning is not lethal, an individual who has been exposed to cyanide is given various nitrites that convert oxyhemoglobin to methemoglobin, which merely involves converting the Fe^{2+} of hemoglobin to Fe^{3+} in methemoglobin. Methemoglobin (Fe^{3+}) competes with cytochrome a,a_3 (Fe^{3+}) for cyanide, forming a methemoglobin cyanide complex. Administration of

(Figure 6.37) inhibit the electron transfer chain at the level of the flavoprotein, NADH dehydrogenase. Hence electrons or reducing equivalents derived from NAD^+-linked dehydrogenases are not oxidized by a rotenone-inhibited respiratory chain, whereas those derived from flavin-linked dehydrogenases are freely oxidized. The antibiotic antimycin A (Figure 6.37) inhibits electron transfer at the level of cytochrone b, whereas the terminal step in the respiratory chain catalyzed by the cytochrome oxidase is inhibited by cyanide, azide, or carbon monoxide. (See Clin. Corr. 6.3.) These later three compounds merely combine with the oxidized heme iron (Fe^{3+}) in cytochromes a and a_3 in order to prevent the reduction of this heme iron by electrons derived from reduced cytochrome c. Hence ingestion or injections of respiratory chain inhibitors leads to a blockage of electron transfer and impairment of the normal energy generating function of the mitochondrial electron transport chain, and if the exposure to such an inhibitor is prolonged, death of the organism would result.

thiosulfate causes the cyanide to react with the enzyme rhodanese forming thiocyanate.



Reversal of Electron Transport

It should be pointed out that the various events in the mitochondrial electron transport system and the closely coupled reactions or pro-

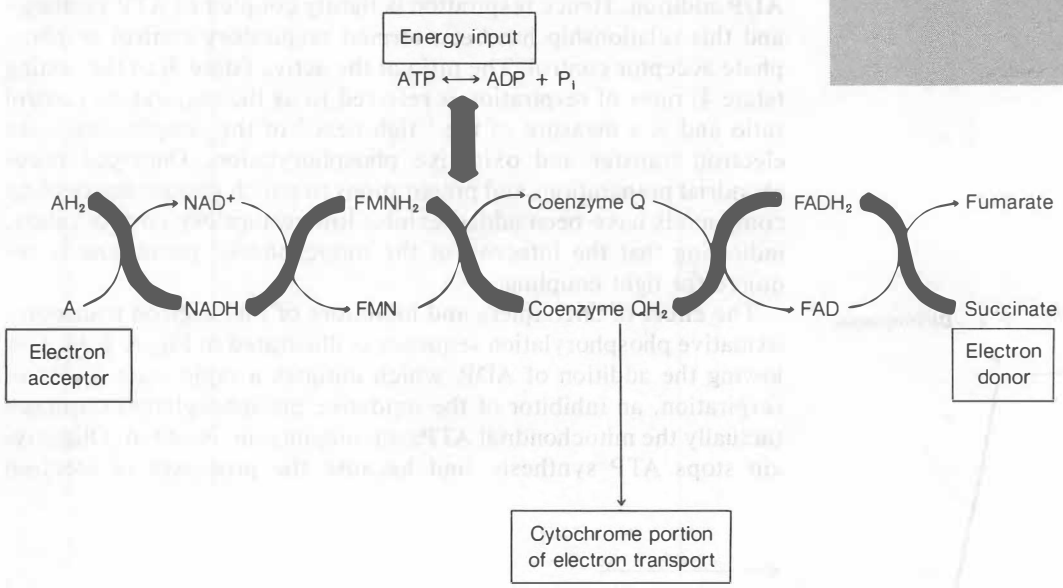


Figure 6.38
Reversal of mitochondrial electron transfer.

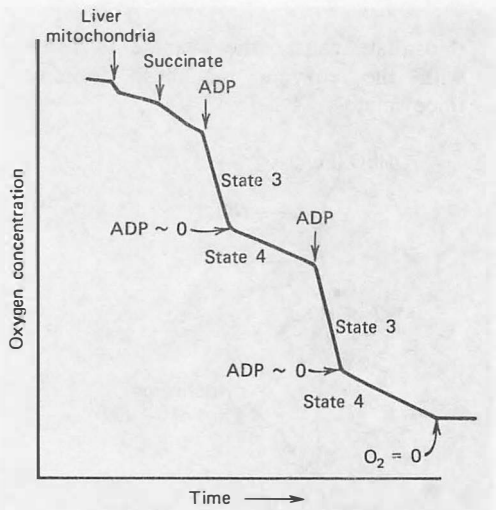
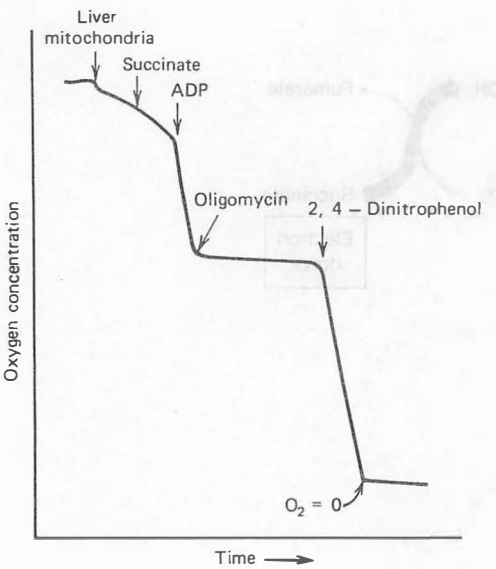


Figure 6.39
A demonstration of the coupling of electron transport to oxidative phosphorylation in a suspension of liver mitochondria.
State 3/state 4 = respiratory control ratio.



cesses in the oxidative phosphorylation sequence are reversible, provided an appropriate amount of energy is supplied to drive the system. It has been observed in mitochondrial systems that reducing equivalents derived from succinate can be transferred to NADH with the concomitant hydrolysis of ATP (see Figure 6.38). Further, it also has been shown that electron transport across the other two phosphorylation sites can be reversed in a similar fashion.

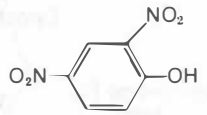
Coupling of Oxidative Phosphorylation to Electron Transport

The obligatory tight coupling between the electron-transferring reactions and the reactions in oxidative phosphorylation can best be illustrated in the experiment shown in Figure 6.39. Mitochondrial electron transport monitored by measuring the rate of oxygen consumption by a suspension of liver mitochondria can occur only at a rapid rate, following the addition of an oxidizable substrate (the electron donor) and ADP (a phosphate acceptor) plus inorganic phosphate. The "active" state in the presence of substrate and ADP has been designated state 3 and is a situation wherein there occurs rapid electron transfer, oxygen consumption, and rapid synthesis of ATP. Following the conversion of all of the added ADP to ATP the rate of electron transfer subsides back to the rate observed prior to ADP addition. Hence respiration is tightly coupled to ATP synthesis and this relationship has been termed respiratory control or phosphate acceptor control. The ratio of the active (state 3) to the resting (state 4) rates of respiration is referred to as the respiratory control ratio and is a measure of the "tightness" of the coupling between electron transfer and oxidative phosphorylation. Damaged mitochondrial preparations and preparations to which various uncoupling compounds have been added exhibit low respiratory control ratios, indicating that the integrity of the mitochondrial membrane is required for tight coupling.

The effect of uncouplers and inhibitors of the electron transport–oxidative phosphorylation sequence is illustrated in Figure 6.40. Following the addition of ADP, which initiates a rapid state 3 rate of respiration, an inhibitor of the oxidative phosphorylation sequence (actually the mitochondrial ATPase), oligomycin, is added. Oligomycin stops ATP synthesis, and because the processes of electron

Figure 6.40
Inhibition and uncoupling of oxidative phosphorylation in liver mitochondria.

transport and ATP synthesis are coupled tightly, respiration or electron transport is inhibited nearly completely. Following the inhibition of both oxygen consumption and ATP synthesis, the addition of an uncoupler of these two processes such as 2,4-dinitrophenol,



Dinitrophenol

causes a rapid initiation of oxygen consumption. Because respiration or electron transport is now uncoupled from ATP synthesis, electron transport may continue but ATP synthesis may not occur.

Finally, it should be noted that the regulation of the respiration rate of a tissue by the provision of a phosphate acceptor, ADP, is a normal physiological situation. For example, when a muscle is exercised, ATP is broken down to ADP and P_i, and creatine phosphate is converted to creatine as the high energy phosphate bond is transferred to ATP in the creatine phosphokinase reaction. As ADP accumulates during the muscular activity, respiration or oxygen consumption is activated, and the energy generated in this fashion allows the ATP and creatine phosphate levels to be replenished. (See Clin. Corr. 6.4.)

Microsomal Electron Transport

Whereas the mitochondrial electron transport chain is linked to the synthesis of ATP in the oxidative phosphorylation sequence there is another type of electron transport chain found in the endoplasmic reticulum or the microsomal fraction of the liver and various other tissues (Figure 6.41). This electron transport chain exhibits several important differences which distinguish it from its mitochondrial counterpart. First, microsomal electron transport utilizes NADPH as the substrate or source of reducing equivalents for the initial flavin-linked reaction in this sequence, for example, the NADPH-cytochrome P_{450} reductase. Second, the microsomal system contains a cytochrome that is unique to this membranous system as it is not found in mitochondria. Cytochrome P_{450} was named as such because the reduced form of the cytochrome has an absorption band at 450 nm. During the terminal reaction of this electron transport sequence the Fe^{2+} of cytochrome P_{450} reacts with molecular oxygen to form a cytochrome P_{450} Fe^{2+}–O_2 complex. Thereafter, one of the oxygen atoms is incorporated into the substrate in a hydroxylation reaction, while the other oxygen is reduced to water. This electron transport

CLIN. CORR. 6.4 HYPOXIC INJURY

Acute hypoxic tissue injury has been studied in a variety of human tissues. The occlusion of one of the major coronary arteries during a myocardial infarction produces a large array of biochemical and physiological sequelae. When a tissue is deprived of its oxygen supply, the mitochondrial electron transport–oxidative phosphorylation sequence is inhibited, resulting in the decline of cellular levels of ATP and creatine phosphate. As cellular ATP levels diminish, anaerobic glycolysis is activated in an attempt to maintain normal cellular functions. Glycogen levels are rapidly depleted and lactic acid levels in the cytosol increase, reducing the intracellular pH. Hypoxic cells in such an energetic deficit begin to swell as they can no longer maintain their normal intracellular ionic environments. Mitochondria swell and begin to accumulate calcium, which may be deposited in the matrix compartment as calcium phosphate. The cell membranes of swollen cells become more permeable, leading to the leakage of various soluble enzymes, coenzymes, and other cell constituents from the cell. As the intracellular pH falls, damage occurs to lysosomal membranes, which release various hydrolytic proteases, lipases, glucosidases, and phosphatases into the cell. Such lysosomal enzymes begin an autolytic digestion of cellular components.

Cells that have been exposed to short periods of hypoxia can recover, without irreversible damage, upon reperfusion with an oxygen-containing medium. The exact point at which hypoxic cell damage becomes irreversible is not precisely known.

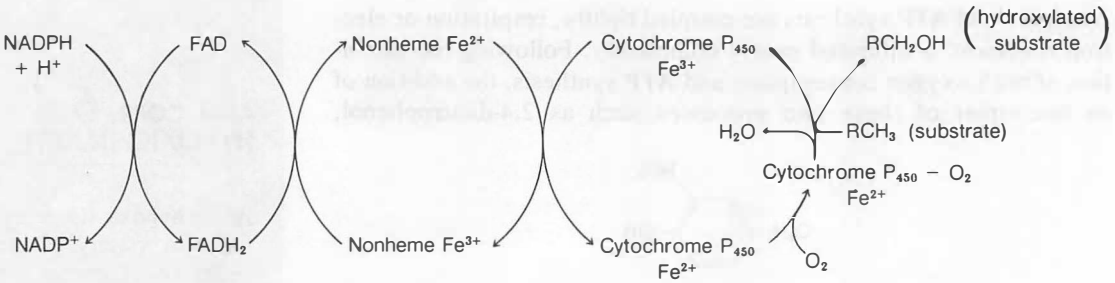


Figure 6.41
Microsomal electron transport.

chain is nonphosphorylating, in that ATP is not synthesized during the transfer of electrons. The primary purpose of this system is in the hydroxylation of various drugs (e.g., phenobarbital), steroids or sterols, fatty acids, polycyclic hydrocarbons, and some amino acids. In fact the synthesis of components in the microsomal electron transport sequence is induced in the liver upon administration of certain drugs or other substrates of this system.

6.7 OXIDATIVE PHOSPHORYLATION

One of the most vexing problems that has confronted biochemists during the past three decades is the delineation of the mechanism of oxidative phosphorylation. Despite the countless man-years of experimental consideration that have been expended on this problem, a precise description of the mechanism by which energy derived from the passage of electrons sequentially along the electron transport chain is transduced into the chemical energy involved in the phosphoanhydride bonds of ATP is not available. Many hypotheses for the mechanism of oxidative phosphorylation have been tested, three general theories have emerged as reasonable proposals, and one of these theories is now widely accepted as the probable mechanism for this process.

Chemical Coupling Hypothesis

The chemical-coupling hypothesis for oxidative phosphorylation was developed initially in the early 1950s. This mechanism was

based upon an analogy with the mechanism for substrate level phosphorylation observed in the glyceraldehyde 3-phosphate dehydrogenase reaction of the glycolytic pathway (Figure 6.42). In this reaction glyceraldehyde 3-phosphate is oxidized and a high energy phosphoric–carboxylic acid anhydride bond is generated in the product of the reaction, 1,3-diphosphoglycerate. An enzyme-bound high energy intermediate is generated in this reaction, which is utilized to form the intermediate high energy compound 1,3-diphosphoglycerate and ultimately to form ATP in the next reaction in the glycolytic pathway, phosphoglycerate kinase (see Chapter 7). Another example of a substrate level phosphorylation reaction, which was defined in the 1960s, is the succinyl CoA synthetase reaction of the Krebs cycle. In this reaction the high energy character of succinyl CoA is converted to the phosphoric acid anhydride bond in GTP with the intermediate participation of a high energy, phosphorylated histidine moiety on the enzyme. Originally it was thought (incorrectly) that this phosphohistidine was a high energy intermediate in the oxidative phosphorylation sequence. Because of these types of substrate level phosphorylation reactions it was proposed that the mechanism of mitochondrial energy transduction involved a series of high energy intermediates which were generated

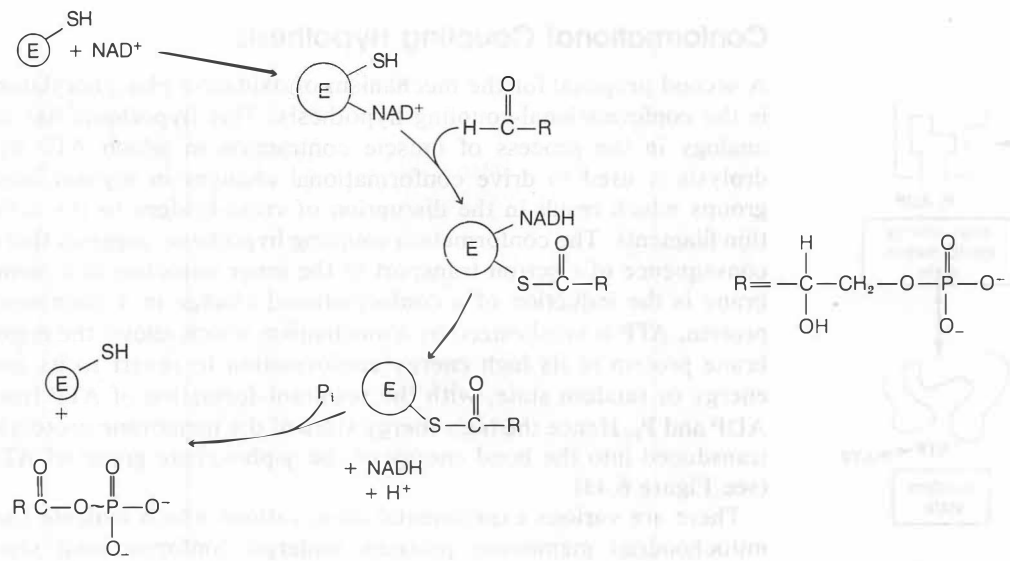


Figure 6.42
The mechanism of the glyceraldehyde 3-phosphate dehydrogenase reaction.

in the mitochondrial membrane as a consequence of electron transport:

$$AH_2 + B + I \longrightarrow A \sim I + BH_2$$

$$A \sim I + P_i \longrightarrow I \sim P_i + A$$

$$I \sim P_i + ADP \longrightarrow ATP + I$$

In this representation A and B are electron carriers, while I is a hypothetical ligand which participates in the formation initially of a high energy compound with the respiratory carrier ($A \sim I$) and thereafter with inorganic phosphate to form a phosphorylated high energy intermediate ($I \sim P_i$). The phosphorylated high energy intermediate is then utilized to form ATP in the ATP synthetase reaction. Uncouplers of oxidative phosphorylation were proposed to act by hydrolyzing the nonphosphorylated high energy intermediate prior to the incorporation of phosphate into the system. Oligomycin was suggested to inhibit the incorporation of phosphate into ATP. The strongest argument for this type of mechanism was its basic simplicity, while its primary detraction is the fact that none of the proposed high energy intermediates have ever been defined or isolated. Hence the suggestion has been made that such intermediates do not actually exist.

Conformational Coupling Hypothesis

A second proposal for the mechanism of oxidative phosphorylation is the conformational-coupling hypothesis. This hypothesis has an analogy in the process of muscle contraction in which ATP hydrolysis is used to drive conformational changes in myosin head groups which result in the disruption of cross-bridges to the actin thin filaments. The conformation coupling hypothesis suggests that a consequence of electron transport in the inner mitochondrial membrane is the induction of a conformational change in a membrane protein. ATP is synthesized by a mechanism which allows the membrane protein in its high energy conformation to revert to its low energy or random state, with the resultant formation of ATP from ADP and P_i. Hence the high energy state of the membrane protein is transduced into the bond energy of the γ-phosphate group of ATP (see Figure 6.43).

There are various experimental observations which indicate that mitochondrial membrane proteins undergo conformational state changes during the process of active electron transport. However, there is relatively little evidence demonstrating conclusively that

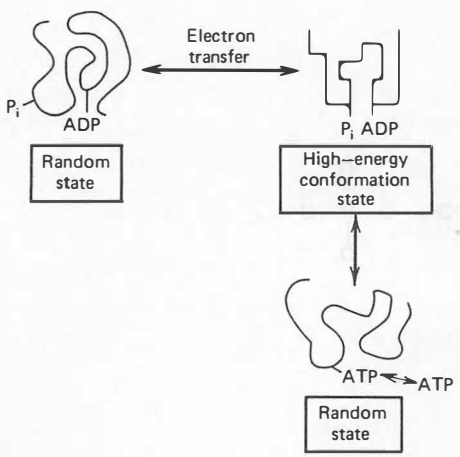


Figure 6.43
The conformational coupling hypothesis.

such conformational state changes are actually involved in the mechanism of ATP synthesis.

Chemiosmotic Coupling Hypothesis

Finally, the chemiosmotic-coupling hypothesis originally proposed by Peter Mitchell has gained widespread appreciation as a mechanism for energy transduction in mitochondria, as well as other biological systems. Mitchell's original proposition of the chemiosmotic theory of oxidative phosphorylation compared the energy generating systems in biological membranes to a common storage battery. Just as energy may be stored in batteries because of the separation of positive and negative charges in the different components of the battery, energy may be generated as a consequence of the separation of charges in complex membranous systems. The chemiosmotic hypothesis (Figure 6.44) suggests that an electrochemical or proton gradient is established across the inner mitochondrial membrane during electron transport. This proton gradient is formed by pumping protons from the mitochondrial matrix side of the inner membrane to the cytosolic side of the membrane. Once there is a substantial electrochemical gradient established, the subsequent dissipation of the gradient is coupled to the synthesis of ATP by the mitochondrial

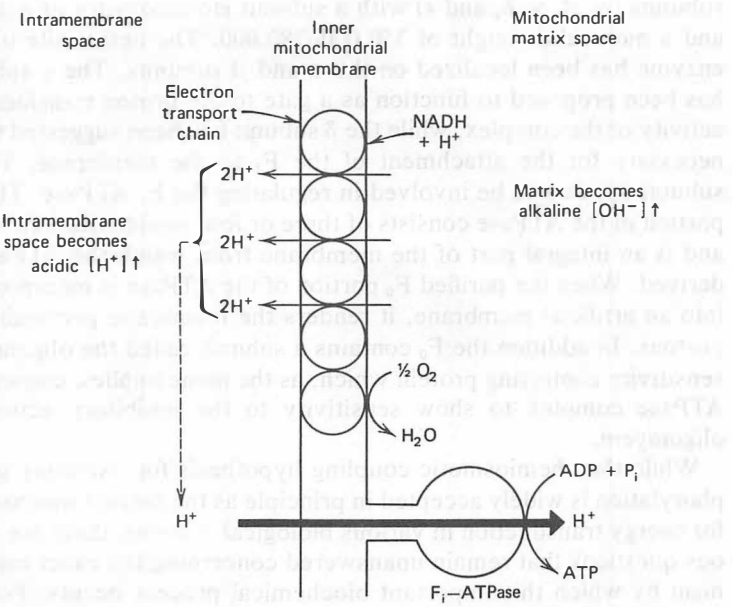


Figure 6.44
The chemiosmatic coupling hypothesis.

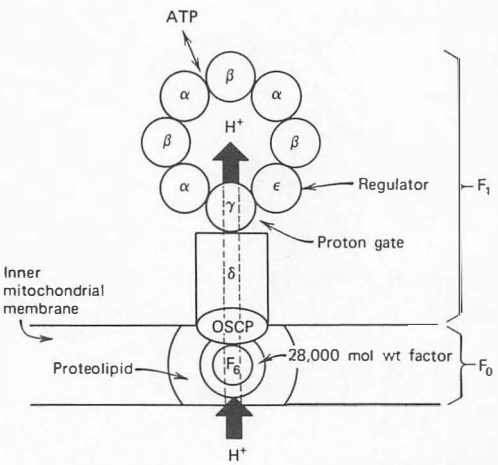


Figure 6.45
A model for the mitochondrial ATPase.

ATPase. The chemiosmotic hypothesis requires that the electron transport carriers and the F_1-ATPase are localized in such a fashion in the inner mitochondrial membrane that protons are pumped out of the matrix compartment during the electron transport phase of the process, and protons are pumped or allowed back through the membrane during the ATP synthetase aspect of the process.

Uncouplers, which are usually relatively lipophilic weak acids, act to dissipate the proton gradient by transporting protons through the membrane from the intramembrane space to the matrix, essentially short-circuiting the normal flow of protons through the ATP synthetic portion of the system. One of the strongest arguments supporting the chemiosmotic hypothesis is that soluble ATPases have been purified, incorporated into artificial membrane vesicles, and are able to synthesize ATP when an electrochemical gradient is established across the artificial membrane. In recent years a considerable experimental effort has been expended to purify the various components of the mitochondrial ATPase. It has been determined that proton-translocating ATPases are present and may be purified from a variety of mammalian tissues, bacteria, and yeast. The ATPase is a multicomponent complex with a suggested molecular weight of 480,000–500,000 (Figure 6.45). These ATPases can be incorporated into artificial membranes and can catalyze ATP synthesis. The ATPase complex consists of a water-soluble portion called F_1 and a hydrophobic portion called F_0. The F_1 consists of five nonidentical subunits (α, β, γ, δ, and ϵ) with a subunit stoichiometry of $\alpha_3\beta_3\gamma\delta\epsilon$ and a molecular weight of 350,000–380,000. The active site of the enzyme has been localized on the α and β subunits. The γ subunit has been proposed to function as a gate to the proton translocating activity of the complex, while the δ subunit has been suggested to be necessary for the attachment of the F_1 to the membrane. The ϵ subunit appears to be involved in regulating the F_1 ATPase. The F_0 portion of the ATPase consists of three or four nonidentical subunits and is an integral part of the membrane from which the ATPase is derived. When the purified F_0 portion of the ATPase is incorporated into an artificial membrane, it renders the membrane permeable to protons. In addition the F_0 contains a subunit called the oligomycin sensitivity conferring protein which, as the name implies, causes the ATPase complex to show sensitivity to the inhibitory action of oligomycin.

While the chemiosmotic-coupling hypothesis for oxidative phosphorylation is widely accepted in principle as the correct mechanism for energy transduction in various biological systems, there are various questions that remain unanswered concerning the exact mechanism by which this important biochemical process occurs. For instance, what is the mechanism by which protons are pumped out of

the mitochondrial matrix during electron transport? What is the stoichiometry of protons pumped per ATP synthesized? What is the mechanism by which protons are pumped back into the matrix "through" the F_1-ATPase? Is there a high energy intermediate involved at some point in the ATP synthetic process?

BIBLIOGRAPHY

Energy-Producing and Energy-Utilizing Systems

Atkinson, D. E. *Cellular energy metabolism and its regulation*. New York: Academic, 1972.

Bock, R. M. Adenine nucleotides and properties of pyrophosphate compounds, in P. D. Boyer, H. Lardy, and K. Myrback (eds.), *The enzymes*, 2nd ed., vol. 2. New York: Academic, 1960, pp. 3–38.

Lipmann, F. Metabolic generation and utilization of phosphate bound energy, *Advan. Enzymol.* 1:99, 1941.

Sources and Fates of Acetyl CoA

Denton, R. M., and Halestrap, A. Regulation of pyruvate metabolism in mammalian tissues. *Essays Biochem.* 15:37, 1979.

Reed, L. J. Multienzyme complexes. *Acct. Chem. Res.* 7:40, 1974.

Reed, L. J., Pettit, F. H., and Yeaman, S. J. Pyruvate dehydrogenase complex: structure, function, and regulation, in P. A. Srere and R. W. Estabrook (eds.), *Microenvironments and metabolic compartmentation*. New York: Academic, 1978, pp. 305–315.

The Tricarboxylic Acid Cycle

Hansford, R. G. Control of mitochondrial substrate oxidation. *Curr. Topics Bioenerg.* 10:217, 1980.

Krebs, H. A. The history of the tricarboxylic acid cycle, *Perspect. Biol. Med.* 14:154, 1980.

Lowenstein, J. M. (ed.). *Citric acid cycle: control and compartmentation*. New York: Marcel Dekker, 1969.

Srere, P. M. The enzymology of the formation and breakdown of citrate. *Advan. Enzymol.* 43:57, 1975.

Structure and Compartmentation of Mitochondrial Membranes

Bygrave, F. L. Mitochondria and the control of intracellular calcium. *Biol. Rev.* 53:43, 1978.

Klingenberg, M. The ATP–ADP carrier in mitochondrial membranes, in A. N. Martinosi (ed.), *The enzymes of biological membranes.* New York: Plenum, 1976, pp. 383–438.

LaNoue, K. F., and Schoolwerth, A. C. Metabolite transport in mitochondria. *Annu. Rev. Biochem.* 48:871, 1979.

Williamson, J. R. The role of anion transport in the regulation of metabolism, in R. W. Hanson and M. A. Mehlman (eds.), *Gluconeogenesis: its regulation in mammalian species.* New York: Wiley-Interscience, 1976, pp. 165–220.

Electron Transfer

Baltsheffsky, H., and Baltsheffsky, M. Electron transport phosphorylation. *Annu. Rev. Biochem.* 43:871, 1974.

Chance, B. The nature of electron transfer and energy coupling reactions, *FEBS Lett.* 23:3, 1972.

Oxidative Phosphorylation

Boyer, P. D., Chance, B., Ernster, L., Mitchell, P., Racker, E., and Slater, E. C. Oxidative phosphorylation and photophosphorylation. *Annu. Rev. Biochem.* 46:955, 1977.

Fillingame, R. H. The proton-translocating pumps of oxidative phosphorylation. *Annu. Rev. Biochem.* 49:1079, 1980.

Kagawa, Y., Sone, N., Hirata, H., and Yooshida, M. Structure and function of H^+ ATPases. *J. Bioenerget. Biomemb.* 11:39, 1979.

Mitchell, P. Keilin's respiratory chain concept and its chemiosmotic consequences. *Science* 206:1148, 1979.

Racker, E. From Pasteur to Mitchell: a hundred years of bioenergetics. *Fed. Proc.* 39:210, 1980.

7

Carbohydrate Metabolism I

ROBERT A. HARRIS

MAJOR METABOLIC PATHWAYS AND THEIR CONTROL

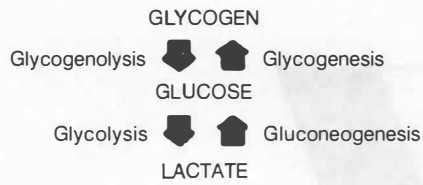


Figure 7.1
Relationship of glucose to the major pathways of carbohydrate metabolism.

7.1 OVERVIEW

The major pathways of carbohydrate metabolism either begin or end with glucose (Figure 7.1). This chapter will describe the utilization of glucose as a source of energy, the formation of glucose from noncarbohydrate precursors, the storage of glucose in the form of glycogen for later use, and the subsequent release of glucose from this storage form for use by cells. A thorough understanding of the pathways and the details of their metabolic regulation is necessary because of the important role played by glucose in the body. Glucose is the major form in which the carbohydrate coming from the intestinal tract is presented to the cells of the rest of the body. Glucose is the only fuel used to any significant extent by a few specialized cells, and it is the major fuel used by the brain. Indeed, glucose is so important to these specialized cells and the brain that several of the major tissues of the body work together to ensure a continuous supply of this essential substrate. Of importance to the practicing physician, glucose metabolism is defective in two very common metabolic diseases, obesity and diabetes, which in turn are contributing factors in the development of a number of major medical problems, including atherosclerosis, hypertension, small vessel diseases, kidney diseases, and blindness.

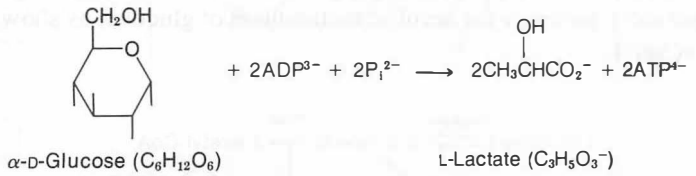
The discussion begins with glycolysis, a pathway that can be used by all cells of the body to extract part of the chemical energy inherent in the glucose molecule. This pathway also converts glucose to pyruvate and thus sets the stage for the complete oxidation of glucose to CO_2 and H_2O. The de novo synthesis of glucose, that is, gluconeogenesis, is considered next. It is a function of the liver and kidneys and can be conveniently discussed following glycolysis because gluconeogenesis will seem, without careful examination of the pathway, to be simply the reverse of the glycolytic pathway. In contrast to glycolysis, which produces ATP, gluconeogenesis requires ATP and is therefore an energy-requiring process. The consequence is that only some of the enzyme-catalyzed steps can be common to both the glycolytic and gluconeogenic pathways. Indeed, the mitochondrion and additional enzyme-catalyzed steps necessarily become involved to make the overall process of gluconeogenesis exergonic. (Note on a confusing point: Gluconeogenesis is an energy-requiring process, that is, it requires ATP, but in order to occur the overall process has to be exergonic and is exergonic because of the ATP-driven steps.) Regulation of the rate-limiting and key enzyme-catalyzed steps will be stressed throughout the chapter. This will be particularly true for glycogen synthesis (glycogenesis) and glycogen degradation (glycogenolysis). Many cells store glyco-

gen for the purpose of having glucose on board for later use. The liver is less selfish, storing glycogen not for its own use, but rather for the maintenance of blood glucose levels to help ensure that other tissues of the body, especially the brain, have an adequate supply of this important substrate. Regulation of the synthesis and degradation of glycogen has been extensively studied and now serves as a model for our current understanding of how hormones work and how other metabolic pathways may be regulated. This subject will be emphasized because it contributes much to our understanding of the diabetic condition, starvation, and how tissues of the body respond to stress, severe trauma, and injury.

7.2 GLYCOLYSIS

The Importance of the Glycolytic Pathway

The Embden–Meyerhof or glycolytic pathway represents an ancient process, possessed by all cells of the human body, in which anaerobic degradation of glucose to lactate occurs. This is one example of anaerobic fermentation, a term used to refer to pathways by which organisms extract chemical energy from high energy fuels in the absence of molecular oxygen. Glycolysis represents an emergency energy-yielding pathway, capable of yielding 2 moles of ATP from a mole of glucose in the absence of molecular oxygen. This means that when the oxygen supply is shut off to a tissue, ATP levels can still be maintained for at least a short period of time by glycolysis. Many examples could be given, but the capacity to turn to glycolysis as a source of energy is particularly important to the human being at birth. With the exception of the brain, circulation of blood decreases to most parts of the body of the neonate during delivery. The brain is not normally deprived of oxygen during delivery, but other tissues must depend upon glycolysis for their supply of ATP until circulation returns to normal and oxygen becomes available once again. This conserves oxygen for use by the brain,

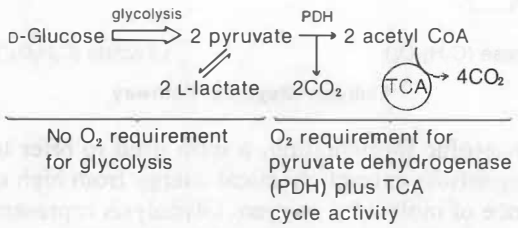


Embden–Meyerhof Pathway

illustrating one of many mechanisms that have evolved to assure survival of brain tissue in times of stress. Glycolysis also sets the stage for aerobic oxidation of carbohydrate in cells. Oxygen is not necessary for glycolysis, and the presence of oxygen can indirectly suppress glycolysis, a phenomenon called the Pasteur effect, that is considered in a later section. Nevertheless, glycolysis can and does occur in cells with an abundant supply of molecular oxygen. Provided that the cells also contain mitochondria, the end product of glycolysis in the presence of oxygen becomes pyruvate rather than lactate. Pyruvate can then be completely oxidized to CO_2 and H_2O by enzymes housed within the mitochondria. The overall process of glycolysis *plus* the subsequent mitochondrial processing of pyruvate to CO_2 and H_2O has the following equation:

$$\text{D-glucose } (C_6H_{12}O_6) + 6O_2 + 38ADP^{3-} + 38P_i^{2-} + 38H^+ \longrightarrow 6CO_2 + 6H_2O + 38ATP^{4-}$$

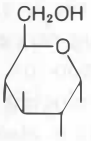
It should be noted that much more ATP is produced in the complete oxidation of glucose to CO_2 and H_2O than in the conversion of glucose to lactate. This has important consequences, which are considered in detail later. The important point at the moment is that, in order for glucose to be completely oxidized to CO_2 and H_2O, it must first be converted to pyruvate by glycolysis. This makes glycolysis a preparatory pathway for aerobic metabolism of glucose, as shown in the scheme.



The importance of glycolysis as a preparatory pathway is best exemplified by the brain. This tissue has an absolute need for glucose and processes most of it via the glycolytic pathway. The pyruvate obtained is then oxidized completely to CO_2 and H_2O in brain mitochondria. Approximately 120 g of glucose is used by the adult human brain each day in order to meet its extraordinary need for ATP. The brain makes extensive use of glycolysis as a means of "preparing" the carbon of glucose for complete oxidation. In contrast, glycolysis per se, with lactate as the end product, is the major mechanism of ATP production in a number of other tissues. Red

blood cells (erythrocytes) lack mitochondria and therefore are unable to convert pyruvate to CO_2 and H_2O. The cornea and lens of the eye have a limited blood supply and also lack mitochondria (because mitochondria would absorb and scatter light) and likewise depend on glycolysis as the major mechanism for ATP production. Kidney medulla, testis, and white muscle fibers are almost totally dependent upon glycolysis as a source of ATP, again because these tissues have relatively few mitochondria. Combined, the tissues that are dependent primarily upon glycolysis for ATP production consume about 40 g of glucose per day in the normal human adult.

In any listing of the importance of glycolysis, it is impossible to ignore alcoholic fermentation. The overall balanced equation for the most common type is given here. This pathway plays an important



$$+ 2ADP^{3-} + 2P_i^{2-} + 2H^+ \longrightarrow 2CH_3CH_2OH + 2CO_2 + 2ATP^{4-}$$

α-D-Glucose ($C_6H_{12}O_6$) Ethanol

role in the making of "good brews," one of those things that make life worth living. This pathway is found in many yeast and certain bacteria but, somewhat surprisingly, the human body also produces a significant amount of ethanol. Most of the production, if not all, may be accounted for by microorganisms present in the intestinal tract. (There is a theory, however, that some of us have a better intestinal flora for the production of ethanol than others!) The pathway of alcoholic fermentation involves the same enzyme-catalyzed reactions as the glycolytic pathway, with the exception that lactate is not the end product. Rather than being reduced to lactate, pyruvate is decarboxylated to give acetaldehyde, which is reduced to ethanol to complete the pathway.

The major dietary sources of glucose are indicated in Chapter 24 in the discussion of the enzymes of digestion. Recall that starch is the storage form of glucose in plants and that it contains α-[1 → 4] glycosidic linkages along with α-[1 → 6] branches. Glycogen is the storage form of glucose in animal tissues and contains the same sort of glycosidic linkages and branches. It is important to distinguish between endogenous and exogenous sources of glucose. Exogenous refers to that which we eat and digest in the intestinal tract, whereas endogenous refers to that which is stored or synthesized in our tissues. Exogenous starch or glycogen is hydrolyzed within the lumen of the intestinal tract with the production of glucose, whereas stored glycogen endogenous to our tissues is converted to glucose or glu-

cose 6-phosphate by enzymes present within the cells of these tissues. The disaccharides, which are important sources of glucose in our diet, include milk sugar (lactose) and grocery store sugar (sucrose). The hydrolysis of these sugars by enzymes of the brush border of the intestinal tract is discussed on page 1159. Glucose can be used as a source of energy to satisfy the needs of the cells of the intestinal tract. However, these cells are not designed to depend upon glucose to any great extent, most of their energy requirement being met by glutamine catabolism. Most of the glucose passes through the cells of the intestinal tract into the blood, where it goes by way of the portal blood and the general circulation to be used by other cells of the body. The first major tissue to have an opportunity to remove it from the portal blood is the liver. This is strategically important because the liver is the major organ involved in buffering blood glucose levels. When blood glucose is too high, the liver removes glucose from the blood by the glucose-consuming processes of glycogenesis and glycolysis. When blood glucose is too low, the liver supplies the blood with glucose by the glucose-producing processes of glycogenolysis and gluconeogenesis. The liver is also the first organ exposed to the blood flowing from the pancreas and therefore "sees" the highest concentrations of the hormones released from this tissue—glucagon and insulin. These are important hormonal regulators of blood glucose levels, in part because of their regulatory effects upon enzyme-catalyzed steps in the liver.

Overview of What Happens to Glucose in Various Cells

After penetrating the plasma membrane by facilitated diffusion, glucose is metabolized mainly by glycolysis in red blood cells (Figure 7.2*A*). Since red blood cells lack mitochondria, the end product of glycolysis is lactate, which is released from the cells back into the blood plasma. Glucose used by the pentose phosphate pathway (see Chapter 8) in red blood cells provides NADPH, necessary in these cells primarily to keep glutathione in the reduced state. Reduced glutathione, in turn, keeps the sulfhydryl groups of enzymes and membrane proteins in the reduced state. The brain, like red blood cells, takes up glucose by facilitated diffusion in an insulin-independent manner (Figure 7.2*B*). Glycolysis in the brain yields pyruvate, which is then oxidized completely to CO_2 and H_2O, as discussed above. The pentose phosphate pathway is also quite active in these cells, generating part of the NADPH needed for reductive synthesis and the maintenance of protein sulfhydryl groups in the reduced state. Muscle and heart cells readily utilize glucose (Figure 7.2*C*), and transport of glucose into both is dependent upon

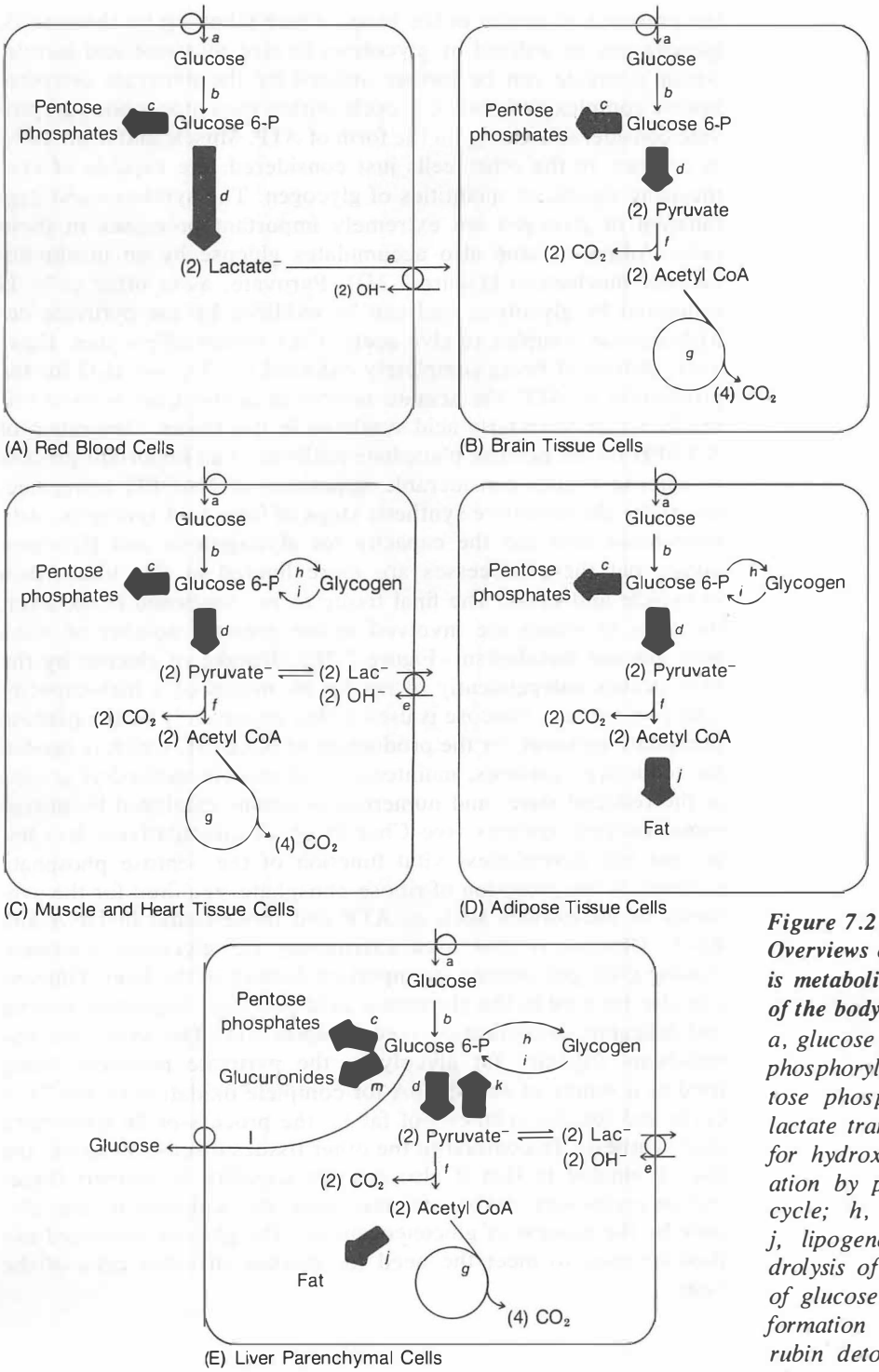


Figure 7.2
Overviews of the major ways in which glucose is metabolized within cells of selected tissues of the body.
a, glucose transport into the cell; b, glucose phosphorylation by hexokinase; c, the pentose phosphate pathway; d, glycolysis; e, lactate transport out of the cell in exchange for hydroxide ion; f, pyruvate decarboxylation by pyruvate dehydrogenase; g, TCA cycle; h, glycogenesis; i, glycogenolysis; j, lipogenesis; k, gluconeogenesis; l, hydrolysis of glucose 6-phosphate and release of glucose from the cell into the blood; m, formation of glucuronides (drug and bilirubin detoxification by conjugation) by the glucuronic acid pathway.

the presence of insulin in the blood. Once taken up by these cells, glucose can be utilized by glycolysis to give pyruvate and lactate. Again pyruvate can be further utilized by the pyruvate dehydrogenase complex and the TCA cycle within the mitochondria to provide considerable energy in the form of ATP. Muscle and heart cells, in contrast to the other cells just considered, are capable of synthesizing significant quantities of glycogen. The synthesis and degradation of glycogen are extremely important processes in these cells. Adipose tissue also accumulates glucose by an insulin-dependent mechanism (Figure 7.2*D*). Pyruvate, as in other cells, is generated by glycolysis and can be oxidized by the pyruvate dehydrogenase complex to give acetyl CoA within adipocytes. However, instead of being completely oxidized to CO_2 and H_2O for the production of ATP, the acetate moiety of acetyl CoA is used primarily for de novo fatty acid synthesis in this tissue. Generation of NADPH by the pentose phosphate pathway is an important process in adipose tissue, considerable quantities of NADPH being necessary for the reductive synthetic steps of fatty acid synthesis. Adipose tissue also has the capacity for glycogenesis and glycogenolysis, but these processes are more limited in this tissue than in muscle and heart. The final tissue to be mentioned is the liver, the cells of which are involved in the greatest number of ways with glucose metabolism (Figure 7.2*E*). Uptake of glucose by the liver occurs independently of insulin by means of a high-capacity transport system. Glucose is used rather extensively by the pentose phosphate pathway for the production of NADPH, which is needed for reductive synthesis, maintenance of protein sulfhydryl groups in the reduced state, and numerous reactions catalyzed by microsomal enzyme systems (see Chapter 6). A quantitatively less important but nevertheless vital function of the pentose phosphate pathway is the provision of ribose phosphate, required for the synthesis of nucleotides such as ATP and those found in DNA and RNA. Glucose is also used extensively for glycogen synthesis, making glycogen storage an important feature of the liver. Glucose can also be used in the glucuronic acid pathway, important in drug and bilirubin detoxification (see Chapter 21). The liver also has significant capacity for glycolysis, the pyruvate produced being used as a source of acetyl CoA for complete oxidation by the TCA cycle and for the synthesis of fat by the process of de novo fatty acid synthesis. In contrast to the other tissues discussed above, the liver is unique in that it also has the capacity to convert three-carbon precursors, such as lactate, pyruvate, and alanine, into glucose by the process of gluconeogenesis. The glucose produced can then be used to meet the need for glucose of other cells of the body.

7.3 THE GLYCOLYTIC PATHWAY

Glucose is combustible and will burn in a test tube to yield heat and light but, of course, no ATP. Cells use some 30 steps to take glucose to CO_2 and H_2O, a seemingly inefficient process, since it can be done in a single step in a test tube. However, side reactions and some of the actual steps used by the cell to "burn" glucose to CO_2 and H_2O lead to the conservation of a significant amount of energy in the form of ATP. In other words, ATP is produced by the controlled burning of glucose in the cell, glycolysis representing only the first few steps, shown in Figure 7.3, in the overall process.

Glycolysis can be conveniently pictured as occurring in three major stages (also see Figure 7.3):

Priming stage:

D-Glucose + $2ATP^{4-} \rightarrow$

D-fructose 1,6-bisphosphate^{4-} + $2ADP^{3-}$ + $2H^+$

Splitting stage:

D-Fructose 1,6-bisphosphate$^{4-} \rightarrow$ 2 D-glyceraldehyde 3-phosphate^{2-}

Oxidoreduction–phosphorylation stage:

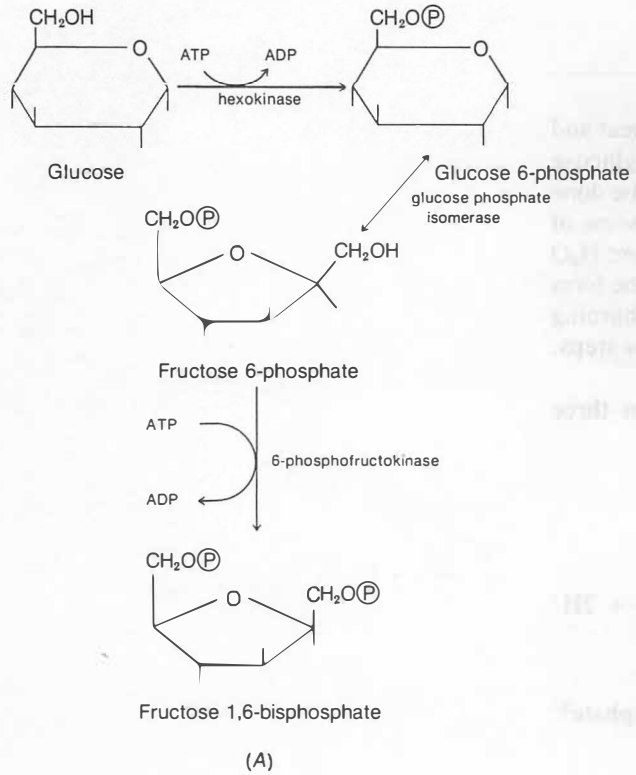
2 D-Glyceraldehyde 3-phosphate^{2-} + $4ADP^{3-}$ + $2P_i^{2-}$ + $2H^+ \rightarrow$

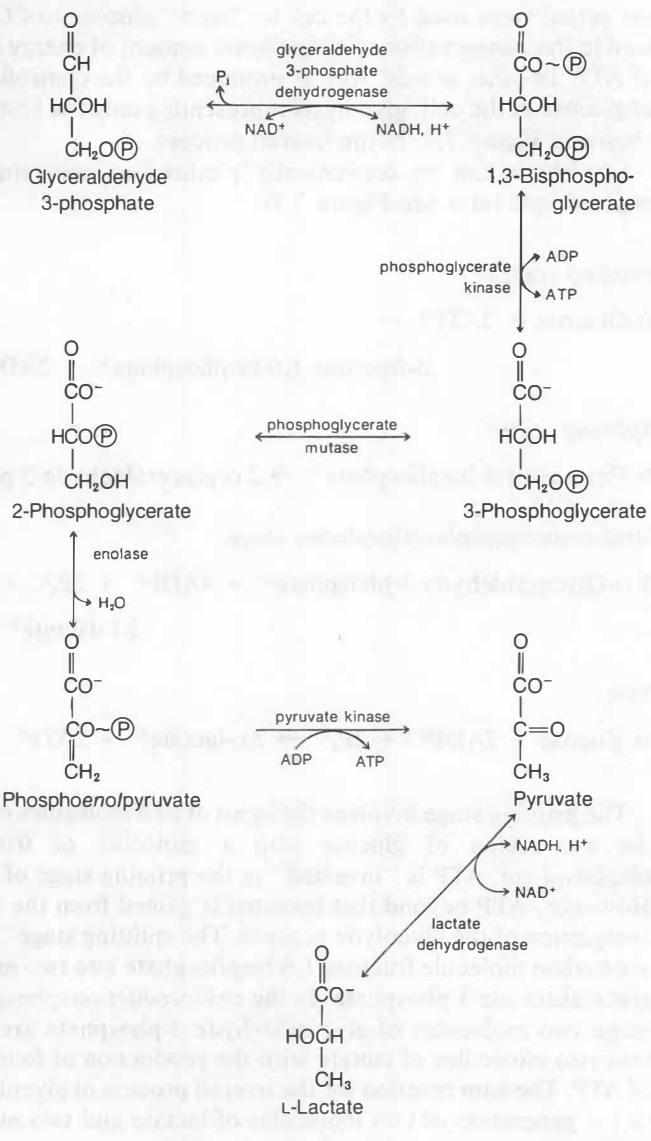
2 L-lactate^{1-} + $4ATP^{4-}$

Sum:

D-glucose + $2ADP^{3-}$ + $2P_i^{2-} \rightarrow$ 2 L-lactate^{1-} + $2ATP^{4-}$

The priming stage involves the input of two molecules of ATP with the conversion of glucose into a molecule of fructose 1,6-bisphosphate. ATP is "invested" in the priming stage of glycolysis. However, ATP beyond that invested is gained from the subsequent completion of the glycolytic process. The splitting stage "splits" the six-carbon molecule fructose 1,6-bisphosphate into two molecules of glyceraldehyde 3-phosphate. In the oxidoreduction–phosphorylation stage two molecules of glyceraldehyde 3-phosphate are converted into two molecules of lactate with the production of four molecules of ATP. The sum reaction for the overall process of glycolysis comes to the generation of two molecules of lactate and two molecules of ATP at the expense of one molecule of glucose.


CH2OH
O
ATP
ADP
hexokinase
CH2OⓅ
Glucose
Glucose 6-phosphate
glucose phosphate isomerase
Fructose 6-phosphate
6-phosphofructokinase
Fructose 1,6-bisphosphate
(A)
or
Fructose 1,6-bisphosphate
fructose-bisphosphate aldolase
triose phosphate isomerase
Dihydroxyacetone phosphate
Glyceraldehyde 3-phosphate
(B)


glyceraldehyde 3-phosphate dehydrogenase
Pi
NAD+
NADH, H+
Glyceraldehyde 3-phosphate
1,3-Bisphospho-glycerate
phosphoglycerate kinase
ADP
ATP
phosphoglycerate mutase
2-Phosphoglycerate
3-Phosphoglycerate
enolase
H2O
Phosphoenolpyruvate
pyruvate kinase
Pyruvate
NADH, H+
NAD+
lactate dehydrogenase
L-Lactate
(C)

Priming Stage

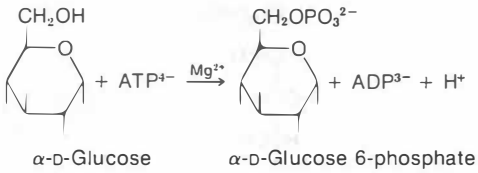
Hexokinase catalyzes the first step of the glycolytic pathway (see Figure 7.3*A* and Step 1). Although this reaction consumes ATP, it gets glycolysis off to a good start by trapping glucose in the form of glucose 6-phosphate within the cytosol of the cell where all of the glycolytic enzymes are located. Phosphate esters of charged, hydrophilic compounds do not readily penetrate cell membranes. The phosphorylation of glucose with ATP is a thermodynamically favorable reaction, requiring the use of one high energy phosphate bond. It is an irreversible reaction under the conditions that exist in cells and represents, therefore, a way to synthesize glucose 6-phosphate. However, it is *not*, by the reverse reaction, a way to synthesize ATP or to hydrolyze glucose 6-phosphate to give glucose. Hydrolysis of glucose 6-phosphate is accomplished by a completely different reaction, catalyzed by the enzyme glucose 6-phosphatase:

$$\text{Glucose 6-phosphate}^{2-} + H_2O \longrightarrow \text{glucose} + P_i^{2-}$$

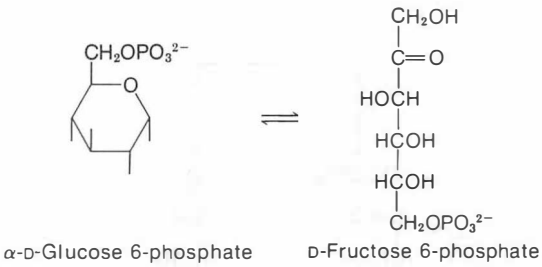
This reaction is thermodynamically favorable in the direction written and cannot be used under conditions existing within biological cells for the synthesis of glucose 6-phosphate from glucose. (A common mistake is to notice that ATP and ADP are involved in the reaction catalyzed by hexokinase but not to notice that they are *not* involved in the reaction catalyzed by glucose 6-phosphatase.) Glucose 6-phosphatase is an important enzyme in liver, functioning to produce free glucose from glucose 6-phosphate in the last step of both gluconeogenesis and glycogenolysis, but it plays no role in the glycolytic pathway.

The next reaction is a readily reversible step of the glycolytic pathway, catalyzed by the enzyme phosphoglucoisomerase (Step 2). This step is not subject to regulation and, since it is readily reversible, functions in both glycolysis and gluconeogenesis.

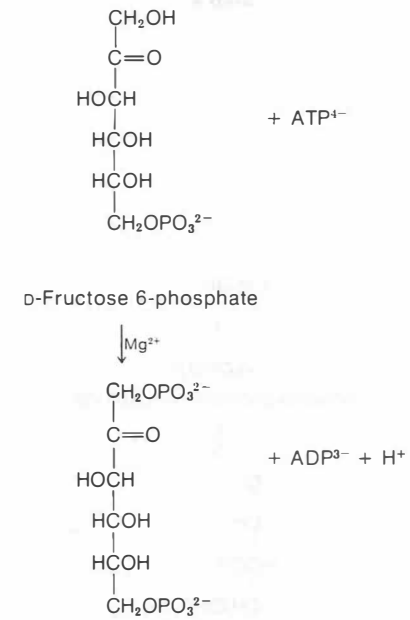
Phosphofructokinase catalyzes the next reaction of the glycolytic pathway, an ATP-dependent phosphorylation of fructose 6-phosphate to give fructose 1,6-bisphosphate (Step 3). This is the



Step 1

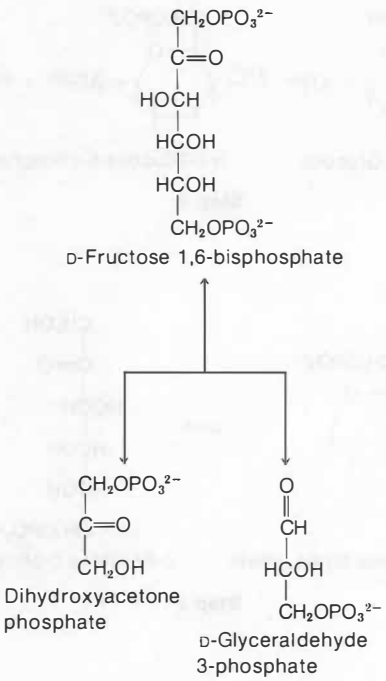


Step 2

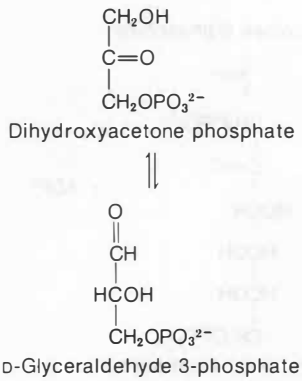


Step 3

Figure 7.3
The glycolytic pathway, divided into its three stages.
The symbol Ⓟ refers to the phosphoryl group PO_3^{2-}; ~ indicates a high energy phosphate bond. (A) Priming stage. (B) Splitting stage. (C) Oxidoreduction-phosphorylation stage.



Step 4



Step 5

favorite enzyme of most students of biochemistry, being subject to regulation by a score of effectors and considered the rate-limiting enzyme of the glycolytic pathway. The reaction is irreversible under intracellular conditions, that is, it represents a way to produce fructose 1,6-bisphosphate but not a way to produce either ATP or fructose 6-phosphate by the reverse reaction. This reaction utilizes the second ATP needed to "prime" glucose, thereby completing the first stage of glycolysis.

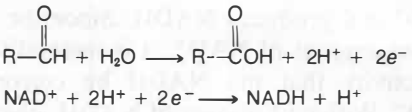
Splitting Stage

Aldolase catalyzes the next step of the glycolytic pathway (see Figure 7.3*B*), cleaving fructose 1,6-bisphosphate into a molecule each of dihydroxyacetone phosphate and glyceraldehyde 3-phosphate (Step 4). This is a reversible reaction, the enzyme being called aldolase because the overall reaction is a variance of an aldol cleavage in one direction and an aldol condensation in the other. Triose phosphate isomerase then catalyzes the reversible interconversion of dihydroxyacetone phosphate and glyceraldehyde 3-phosphate to complete the splitting stage of glycolysis (Step 5). With the transformation of dihydroxyacetone phosphate into glyceraldehyde 3-phosphate, the net conversion of one molecule of glucose into two molecules of glyceraldehyde 3-phosphate has been accomplished.

Oxidoreduction–Phosphorylation Stage

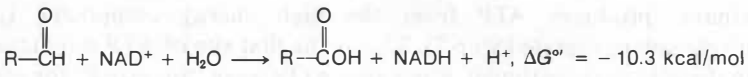
The first reaction of the last stage of glycolysis (Figure 7.3*C*) is catalyzed by the enzyme glyceraldehyde 3-phosphate dehydrogenase (Step 6). This reaction is of considerable interest, not so much because of the regulation of the enzyme involved nor because this complex reaction is reversible under intracellular conditions, but rather because of what is accomplished in a single enzyme-catalyzed step. In this reaction an aldehyde (glyceraldehyde 3-phosphate) is oxidized to a carboxylic acid with the reduction of NAD^+ to NADH. Besides producing NADH, however, the reaction also produces a high energy phosphate compound (1,3-bisphosphoglycerate), which is a mixed anhydride of a carboxylic acid and phosphoric acid. 1,3-Bisphosphoglycerate has a large negative free energy of hydrolysis, enabling it to participate in a subsequent reaction, which yields ATP. The overall reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase can be visualized as the coupling of a very favorable exergonic reaction with a very unfavorable endergonic reaction on the surface of the enzyme. The exergonic reaction can be thought of as being composed of a half-reaction in which an aldehyde (glyceraldehyde 3-phosphate) is oxidized to a carboxylic acid (1,3-

bisphosphoglycerate), which is then coupled with a half-reaction in which NAD^+ is reduced to NADH.

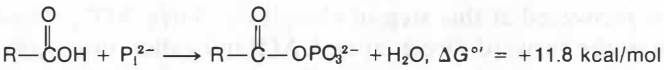


$$NAD^+ + 2H^+ + 2e^- \longrightarrow NADH + H^+$$

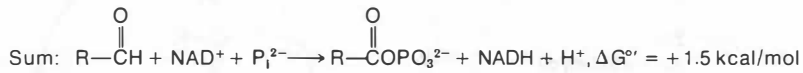
The overall reaction (sum of the half-reactions) is quite exergonic, with the aldehyde being oxidized to a carboxylic acid and NAD^+ being reduced to NADH:



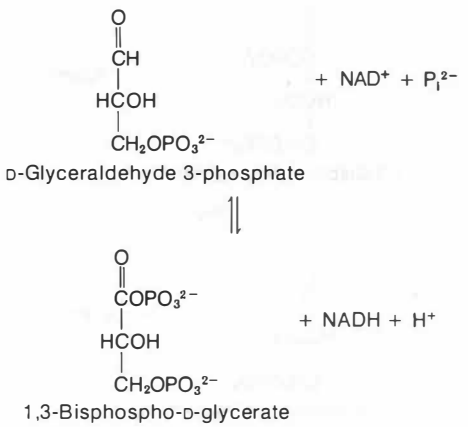
The endergonic component of the reaction corresponds to the formation of a mixed anhydride between the carboxylic acid and phosphoric acid:



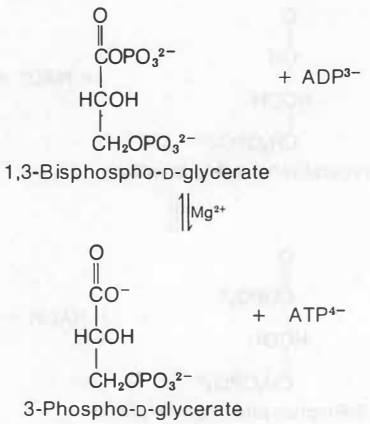
The overall reaction involves coupling of the endergonic and exergonic components to give an overall standard free energy change of +1.5 kcal/mol:



The reaction is freely reversible under intracellular conditions and is used in both the glycolytic and gluconeogenic pathways. The proposed mechanism for the enzyme-catalyzed reaction is shown in Figure 7.4. Glyceraldehyde 3-phosphate reacts with a sulfhydryl group of a cysteine residue of the enzyme to generate a thiohemiacetal. An internal oxidation–reduction reaction takes place on the surface of the enzyme in which the bound NAD^+ is reduced to NADH and the thiohemiacetal is oxidized to give a high energy thiol ester. Exogenous NAD^+ then replaces the bound NADH and the high energy thiol ester reacts with inorganic phosphate to form the mixed anhydride and regenerate the free sulfhydryl group. The mixed anhydride then dissociates from the enzyme. It should be noted, in contrast to the exergonic and endergonic components of the reactions discussed above, that a carboxylic acid (RCO_2H) is not considered to be an intermediate in the actual reaction mechanism. Instead, the enzyme uses the strategy of generating a high energy thiol ester, which can be readily converted into another high energy



Step 6



Step 7

compound, the mixed anhydride between the carboxylic and phosphoric acids.

The reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase requires NAD^+ and produces NADH. Since the cytosol of cells only has a limited amount of NAD^+, it is imperative for continuous glycolytic activity that the NADH be converted back (turned over) to NAD^+. Without turnover of NADH, glycolysis will stop for want of NAD^+. The options which cells have for accomplishing the regeneration of NAD^+ are considered in detail in a later section of this chapter (see page 341).

The next reaction, catalyzed by the enzyme phosphoglycerate kinase, produces ATP from the high energy compound 1,3-bisphosphoglycerate (Step 7). This is the first site of ATP production in the glycolytic pathway. Since two ATPs were "invested" for each glucose molecule in the priming stage [one at the hexokinase-catalyzed step (1) and one at the phosphofructokinase-catalyzed step (3)], and since two molecules of 1,3-bisphosphoglycerate are produced from each glucose, all of the ATP "invested" in the priming stage is recovered at this step of glycolysis. Since ATP production occurs in the forward direction and ATP utilization in the reverse direction, it is somewhat surprising that the reaction is freely rever-

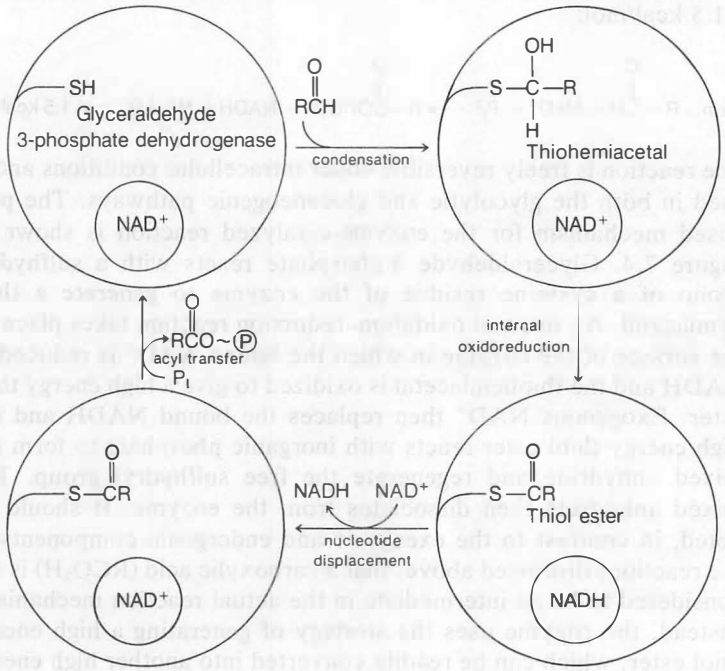
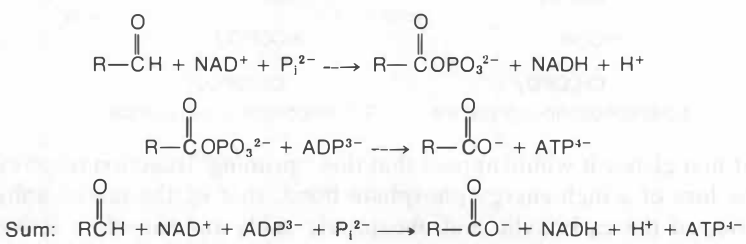
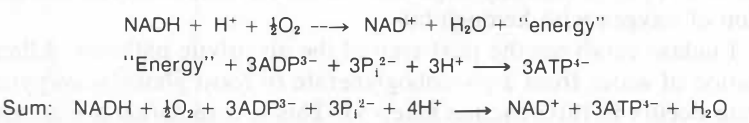


Figure 7.4
The mechanism of action of glyceraldehyde 3-phosphate dehydrogenase.
Large circle represents the enzyme; small circle, the binding site for NAD^+; $R\overset{O}{\overset{\|}{C}}H$, glyceraldehyde 3-phosphate; $-SH$, the sulfhydryl group of the cysteinyl residue located at the active site; and $\sim P$, the high energy phosphate bond of 1,3-bisphosphoglycerate.

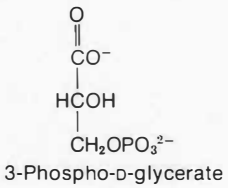
sible and can be used in both the glycolytic and gluconeogenic pathways. The reaction provides a means for the generation of ATP in the glycolytic pathway but, when needed for glucose synthesis, can also be used in the reverse direction for the synthesis of 1,3-bisphosphoglycerate at the expense of ATP. The glyceraldehyde 3-phosphate dehydrogenase-phosphoglycerate kinase system is an example of substrate-level phosphorylation, a term used to refer to a process in which a substrate participates in an enzyme-catalyzed reaction that yields ATP or GTP. Substrate-level phosphorylation stands in contrast to oxidative phosphorylation in which electron transport by the respiratory chain of the mitochondrial inner membrane is used to provide the energy necessary for ATP synthesis (see Chapter 5). Note, however, that the combination of the reactions catalyzed by glyceraldehyde 3-phosphate dehydrogenase and phosphoglycerate kinase accomplishes the coupling of an oxidation (an aldehyde goes to a carboxylic acid) to a phosphorylation:



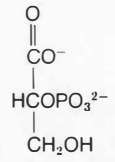
This relationship can be confusing because one needs to be able to differentiate this process from that of oxidative phosphorylation in which the reactions of the mitochondrial respiratory chain are coupled to phosphorylation reactions. Recall that the following reactions correspond to those of oxidative phosphorylation in the mitochondrial inner membrane:



In the case of the glyceraldehyde 3-phosphate dehydrogenase–phosphoglycerate kinase system, an oxidation reaction is well established to precede a substrate-level phosphorylation reaction, but, in contrast to oxidative phosphorylation, no membranes are needed for the action of these enzymes. Although it is conceivable that something comparable takes place within the mitochondrial inner membrane, there is no evidence for reactions of the substrate-

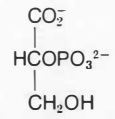


3-Phospho-D-glycerate

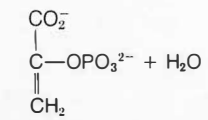


2-Phospho-D-glycerate

Step 8



2-Phospho-D-glycerate



Phosphoenol/pyruvate

Step 9

level phosphorylation type being involved in the process of oxidative phosphorylation.

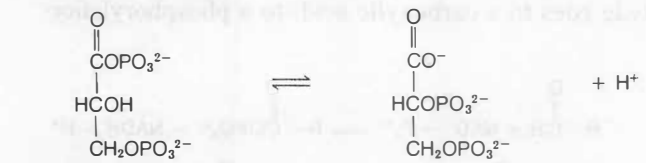
Phosphoglycerate mutase catalyzes the next reaction of the glycolytic pathway, the step in which 3-phosphoglycerate is converted to 2-phosphoglycerate (Step 8). This is a freely reversible reaction in which 2,3-biphosphoglycerate functions as an obligatory intermediate at the active site of the enzyme (E):

$$\text{E + 2,3-bisphosphoglycerate} \rightleftharpoons \text{E-phosphate + 2 phosphoglycerate}$$

$$\text{E-phosphate + 3-phosphoglycerate} \rightleftharpoons \text{E + 2,3-bisphosphoglycerate}$$

$$\text{Sum: 3-Phosphoglycerate} \rightleftharpoons \text{2-phosphoglycerate}$$

2,3-Bisphosphoglycerate is synthesized by a reaction catalyzed by another enzyme, bisphosphoglycerate mutase:



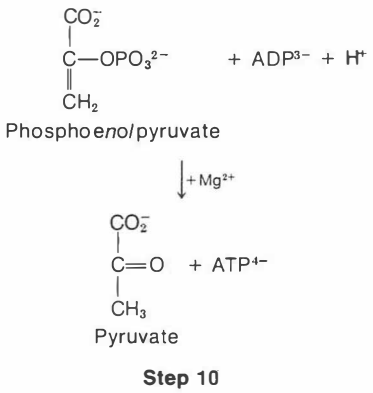
1,3-Bisphospho-D-glycerate 2,3-Bisphospho-D-glycerate

At first glance it would appear that this "priming" reaction results in the loss of a high energy phosphate bond, that is, the mixed anhydride of the carboxylic and phosphoric acids and therefore loss of the capacity to synthesize a molecule of ATP from the high energy bond of 1,3-bisphosphoglycerate. However, the formation of 2,3-bisphosphoglycerate from 1,3-bisphosphoglycerate is reversible, and only catalytic amounts of 2,3-bisphosphoglycerate are needed to serve as an intermediate in the reaction catalyzed by phosphoglycerate mutase. Red blood cells represent a special case, in which 2,3-bisphosphoglycerate accumulates to high concentrations and functions as a physiologically important allosteric effector of the association of oxygen with hemoglobin.

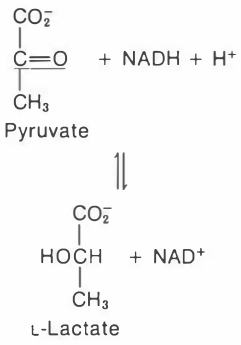
Enolase catalyzes the next step of the glycolytic pathway. Elimination of water from 2-phosphoglycerate to form phospho*enol*pyruvate occurs in this reaction (Step 9). This is a remarkable reaction from the standpoint that a high energy phosphate compound is generated from one of markedly lower energy level. The standard free energy change ($\Delta G^{\circ\prime}$) for the hydrolysis of phospho*enol*pyruvate is -14.8 kcal/mol, a value strikingly greater than that of 2-phosphoglycerate (-4.2 kcal/mol). Although the reaction catalyzed by the enzyme is freely reversible, a large change in the distribution of energy occurs as a consequence of the action of enolase upon 2-phosphoglycerate.

The next step of the glycolytic sequence is catalyzed by pyruvate kinase (Step 10). This enzyme accomplishes the synthesis of ATP with the conversion of the high energy compound phospho*enol*pyruvate into pyruvate. The reaction is not reversible under intracellular conditions. It constitutes a way to synthesize ATP, but in contrast to the reaction catalyzed by phosphoglycerate kinase is not reversible under conditions that exist in cells and is not a reaction that can be used for the synthesis of phospho*enol*pyruvate when needed for glucose synthesis.

The last step of the glycolytic pathway is an oxidoreduction reaction catalyzed by lactate dehydrogenase (Step 11). Pyruvate is reduced in this reaction to give L-lactate, whereas NADH is oxidized to give NAD^+. This is a freely reversible reaction and the only one of the body in which L-lactate participates, that is, the only reaction that can result in L-lactate formation or L-lactate utilization. It should be noted that the NADH generated by glyceraldehyde 3-phosphate dehydrogenase is converted back to NAD^+ by lactate dehydrogenase (see Figure 7.3), the major option used by cells under anaerobic conditions for the regeneration of cytosolic NAD^+.



Step 10



Step 11

Stoichiometries of the Glycolytic Pathway

An examination of the overall glycolytic pathway will show that there is a perfect coupling between the generation of NADH and its utilization (Figure 7.3). Two molecules of NADH are generated at the level of glyceraldehyde 3-phosphate dehydrogenase and two molecules of NADH are utilized at the level of lactate dehydrogenase in the overall conversion of one molecule of glucose into two molecules of lactate. NAD^+, a soluble molecule present in the cytosol, is available in only limited amounts to participate in the glycolytic pathway. It is essential, therefore, that NAD^+ be regenerated from NADH for the glycolytic pathway to continue unabated. The NAD^+ reacts at the level of glyceraldehyde 3-phosphate dehydrogenase to produce NADH, which diffuses through the cytosol until it makes contact with lactate dehydrogenase, which, if a molecule of pyruvate is also available, forms lactate with the regeneration of NAD^+. The overall reaction catalyzed by the combined actions of these two enzymes is the conversion of glyceraldehyde 3-phosphate, pyruvate, and inorganic phosphate into lactate and 1,3-bisphosphoglycerate.

D-Glyceraldehyde 3-phosphate + NAD^+ + P_i ⟶
1,3-bisphospho-D-glycerate + NADH + H^+

Pyruvate + NADH + H^+ ⟶ L-lactate + NAD^+

Sum: D-Glyceraldehyde 3-phosphate + pyruvate + P_i ⟶
1,3-bisphosphoglycerate + L-lactate

This perfect coupling of reducing equivalents in the glycolytic pathway only has to occur under conditions of anaerobiosis, or in cells that lack mitochondria. With the availability of oxygen and mitochondria, reducing equivalents in the form of NADH generated at the level of glyceraldehyde 3-phosphate dehydrogenase can be "shuttled" into the mitochondria for the synthesis of ATP. When this occurs, the end product of glycolysis becomes pyruvate. Two shuttle systems are known to exist for the transport of reducing equivalents from the cytosolic space to the mitochondrial matrix space (mitosol). The mitochondrial inner membrane is not permeable to NADH; therefore, NADH cannot penetrate directly across the mitochondrial inner membrane to gain access to the NADH dehydrogenase of the mitochondrial electron transfer chain.

Shuttle Pathways

The glycerol phosphate shuttle is shown in Figure 7.5*A*; the malate–aspartate shuttle in Figure 7.5*B*. All tissues that have mitochondria appear to also have the capability of "shuttling" reducing equivalents from the cytosol to the mitosol. The relative proportion of the two shuttles varies from tissue to tissue, with liver making greater use of the malate–aspartate shuttle, whereas some muscle cells may be more dependent on the glycerol phosphate shuttle. The shuttle systems are irreversible, that is, they represent mechanisms for moving reducing equivalents into the mitosol, but not mechanisms for moving mitochondrial reducing equivalents into the cytosol.

The transport of aspartate out of the mitochondria in exchange for glutamate is the irreversible step in the malate–aspartate shuttle. The mitochondrial inner membrane has a large number of transport systems (see Chapter 5), but lacks one which is effective for oxalacetate. For this reason oxalacetate transaminates with glutamate to produce aspartate, which then exits irreversibly from the mitochondrion in exchange for glutamate. The aspartate entering the cytosol transaminates with α-ketoglutarate to give oxalacetate and glutamate. The oxalacetate accepts the reducing equivalents of NADH and becomes malate. Malate then penetrates the mitochondrial inner membrane, where it is oxidized by the mitochondrial malate dehydrogenase. This produces NADH within the mitosol and regenerates oxalacetate to complete the cycle. The overall, balanced equation for the sum of all the reactions of the malate–aspartate shuttle is simply,

$$NADH_{cytosol} + H^+_{cytosol} + NAD^+_{mitosol} \longrightarrow NAD^+_{cytosol} + NADH_{mitosol} + H^+_{mitosol}$$

The glycerol phosphate shuttle is simpler, in the sense that fewer reactions are involved, but it should be noted that $FADH_2$ is generated as the end product within the mitochondrial inner membrane, rather than NADH within the mitosolic compartment. The irreversible step of the shuttle is catalyzed by the mitochondrial glycerol 3-phosphate dehydrogenase. The active site of this enzyme is exposed on the cytosolic surface of the mitochondrial inner membrane,

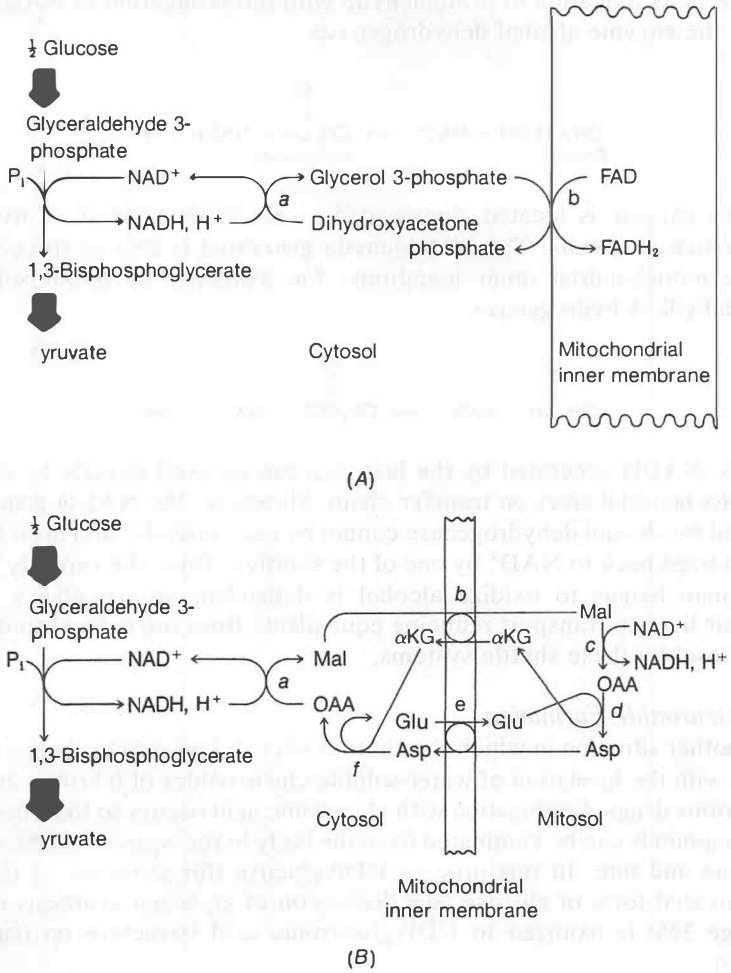


Figure 7.5
The shuttles for the transport of reducing equivalents from the cytosol to the mitochondrial electron-transfer chain.
(A) Glycerol phosphate shuttle: a, cytosolic glycerol 3-phosphate dehydrogenase oxidizes NADH; b, glycerol 3-phosphate dehydrogenase of the outer surface of the mitochondrial inner membrane reduces FAD. (B) Malate–aspartate shuttle: a, cytosolic malate dehydrogenase reduces oxalacetate (OAA) to malate; b, dicarboxylic acid antiport of the mitochondrial inner membrane catalyzes electrically silent exchange of malate for α-ketoglutarate (α-KG); c, mitochondrial malate dehydrogenase produces intramitochondrial NADH; d, mitochondrial aspartate aminotransferase transaminates glutamate and oxalacetate; e, glutamate–aspartate antiport of the mitochondrial inner membrane catalyzes electrogenic exchange of glutamate for aspartate; f, cytosolic aspartate aminotransferase transaminates aspartate and α-ketoglutarate.

making it unnecessary for glycerol 3-phosphate to penetrate completely into the mitosol for oxidation. The overall balanced equation for the sum of the reactions of the glycerol phosphate shuttle is

$$NADH_{cytosol} + H^+ + FAD_{inner\ membrane} \longrightarrow NAD^+_{cytosol} + FADH_{2\ inner\ membrane}$$

Alcohol Oxidation

We should not get trapped into thinking, because of Figure 7.5, that the glycerol phosphate and malate–aspartate shuttles are only designed to handle the NADH generated by glycolysis. These shuttles are important in using NADH generated in the cytosol in other ways as well. For example, the first step of alcohol (i.e., ethanol) metabolism is its oxidation to acetaldehyde with the production of NADH by the enzyme alcohol dehydrogenase.

$$\underset{\text{Ethanol}}{CH_3CH_2OH} + NAD^+ \longrightarrow \underset{\text{Acetaldehyde}}{CH_3\overset{\overset{O}{\|}}{C}H} + NADH + H^+$$

This enzyme is located almost exclusively in the cytosol of liver parenchymal cells. The acetaldehyde generated is able to traverse the mitochondrial inner membrane for oxidation by a mitosolic aldehyde dehydrogenase.

$$CH_3\overset{\overset{O}{\|}}{C}H + NAD^+ \longrightarrow CH_3\overset{\overset{O}{\|}}{C}O^- + NADH + 2H^+$$

The NADH generated by the last step can be used directly by the mitochondrial electron transfer chain. However, the NADH generated by alcohol dehydrogenase cannot be used directly, and must be oxidized back to NAD^+ by one of the shuttles. Thus, the capacity of human beings to oxidize alcohol is dependent on the ability of their liver to transport reducing equivalents from the cytosol to the mitosol by these shuttle systems.

Glucuronide Formation

Another situation in which the shuttles play an important role has to do with the formation of water-soluble glucuronides of bilirubin and various drugs. Conjugation with glucuronic acid occurs so that these compounds can be eliminated from the body in the aqueous media of urine and bile. In this process UDP-glucose (for structure of this activated form of glucose, see discussion of glycogen synthesis on page 386) is oxidized to UDP-glucuronic acid (structure on page 419):

$$\text{UDP-D-glucose} + 2\text{NAD}^+ + \text{H}_3\text{O} \longrightarrow \text{UDP-D-glucuronic acid} + 2\text{NADH} + 2\text{H}^+$$

In a reaction that occurs primarily in the liver, the "activated" glucuronic acid molecule is then transferred to a nonpolar, acceptor molecule, such as some compound (e.g., a drug) foreign to the body:

$$\text{UDP-D-glucuronic acid} + \text{R-OH} \longrightarrow \text{R-}O\text{-glucuronic acid} + \text{UDP}$$

Excess NADH generated by the first reaction has to be eliminated from the cytosol for this process to continue, and, of course, the shuttles are called into play for this purpose. Since ethanol oxidation and drug conjugation are properties of the liver, the two of them occurring together may overwhelm the combined capacity of the shuttles. A good thing to remember—and to tell patients—is not to mix the intake of pharmacologically active compounds and the consumption of alcohol (see Clin. Corr. 7.1).

Energetics of NADH Oxidation

The mitosolic NADH formed as a consequence of malate–aspartate shuttle activity can be used in the presence of oxygen by the mitochondrial respiratory chain for the production of three molecules of ATP by oxidative phosphorylation:

$$\text{NADH}_{\text{mitosol}} + \text{H}^+ + \tfrac{1}{2}\text{O}_2 + 3\text{ADP} + 3\text{P}_\text{i} \longrightarrow \text{NAD}^+_{\text{mitosol}} + 3\text{ATP} + \text{H}_2\text{O}$$

In contrast, the FADH_2 obtained by glycerol phosphate shuttle activity is worth only two ATPs:

$$\text{FADH}_{2\,\text{inner membrane}} + \tfrac{1}{2}\text{O}_2 + 2\text{ADP} + 2\text{P}_\text{i} \longrightarrow \text{FAD}_{\text{inner membrane}} + 2\text{ATP} + \text{H}_2\text{O}$$

Without the intervention of these shuttle systems, the conversion of one molecule of glucose to two molecules of lactate by glycolysis results in the *net* formation of two molecules of ATP. Two molecules of ATP are used in the priming stage to set glucose up so that it can be cleaved. However, subsequent steps then yield four molecules of ATP so that the overall net production of ATP by the glycolytic pathway is two molecules of ATP. Biological cells have only a limited amount of ADP and inorganic phosphate. Flux through the glycolytic pathway is also dependent, therefore, upon an adequate supply of these substrates. Consequently, the ATP generated has to

CLIN. CORR. 7.1 ALCOHOL AND BARBITURATES

Acute alcohol intoxication causes increased sensitivity of an individual to the general depressant effects of barbiturates. This drug combination is very dangerous, normal prescription doses of barbiturates having potentially lethal consequences in the presence of ethanol. In addition to the depressant effects of both ethanol and barbiturates on the central nervous system, ethanol inhibits the metabolism of barbiturates and prolongs the time barbiturates remain effective in the body. Hydroxylation of barbiturates by the endoplasmic reticulum of the liver is inhibited by ethanol. This reaction, catalyzed by the NADPH-dependent cytochrome P_{450} system, results in water soluble derivatives of the barbiturates which are eliminated readily from the circulation by the kidneys. Blood levels of barbiturates remain high when ethanol is present, causing increased central nervous system depression.

Surprisingly, the alcoholic when sober is subsensitive to barbiturates. Chronic ethanol consumption apparently causes adaptive changes in the sensitivity of the central nervous system to barbiturates. It also results in the induction of the enzymes of liver endoplasmic reticulum involved in drug hydroxylation reactions. Consequently, the sober alcoholic is able to metabolize barbiturates more rapidly. This sets up the following scenario. A sober alcoholic has trouble falling asleep, even after taking several sleeping pills, because his liver has increased capacity to hydroxylate the barbiturate contained in the pills. In frustration he consumes more pills and then alcohol. Sleep results, but may be followed by respiratory depression

and death because the alcoholic, although less sensitive to barbiturates when sober, remains sensitive in the presence of alcohol.

be used, that is, turned over, in normal work-related processes in order for glycolysis to occur. The equation for the utilization of ATP for any work-related process is simply

$$ATP^{4-} \longrightarrow ADP^{3-} + P_i^{2-} + H^+ + \text{“work”}$$

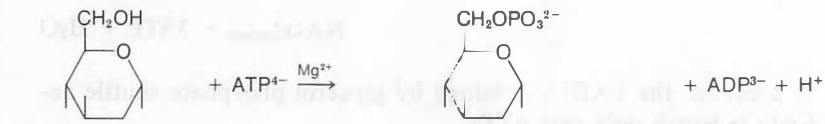
When this equation is added to that given above for glycolysis, excluding the work accomplished, the overall balanced equation for the glycolytic process when coupled to some ATP-utilizing work performance becomes

$$\text{D-Glucose} \longrightarrow 2\ \text{lactate}^- + 2H^+$$

If the ATP is not utilized for work performances, glycolysis will stop for want of ADP and/or inorganic phosphate. Thus glycolytic activity is dependent on the turnover of ATP to ADP and P_i, just as it is dependent on the turnover of NADH to NAD^+.

Poisons of the Glycolytic Pathway

The best known of the poisons of the glycolytic pathway include 2-deoxyglucose, sulfhydryl reagents, and fluoride. 2-Deoxyglucose causes inhibition at the first step, that is, at the reaction catalyzed by hexokinase. 2-Deoxyglucose serves as a substrate for this enzyme, being converted to the 6-phosphate ester by the reaction shown here.

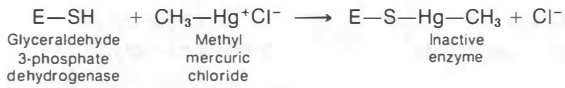


α-D-2-Deoxyglucose α-D-2-Deoxyglucose 6-phosphate

Like glucose 6-phosphate, 2-deoxyglucose 6-phosphate is an effective inhibitor of the reaction catalyzed by hexokinase, but unlike glucose 6-phosphate, 2-deoxyglucose 6-phosphate will not function as a substrate for the reaction catalyzed by phosphoglucoisomerase. Deoxyglucose 6-phosphate inhibition of hexokinase prevents glucose from being phosphorylated, resulting in the inhibition of glycolysis at the very first step.

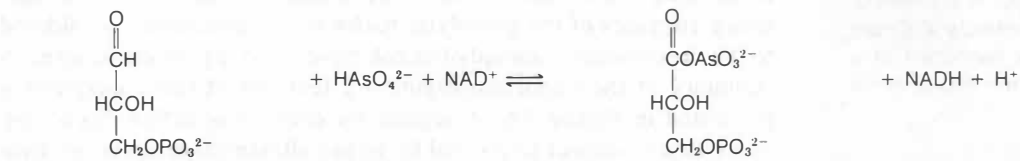
Sulfhydryl reagents bring about an inhibition at the level of glyceraldehyde 3-phosphate dehydrogenase. As discussed above, this enzyme has a cysteinyl residue at the active site, the sulfhydryl group of which reacts with glyceraldehyde 3-phosphate to give a thiohemiacetal. Sulfhydryl reagents are usually mercury-containing compounds or alkylating compounds, such as iodoacetate, which

readily react with the sulfhydryl group of glyceraldehyde 3-phosphate dehydrogenase to prevent the formation of the thiohemiacetal.

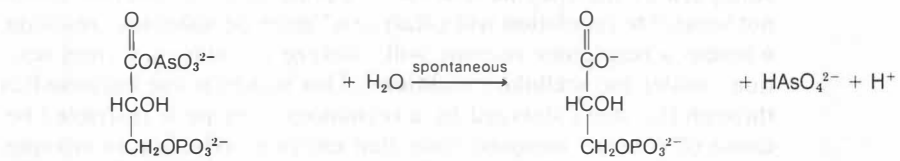


Fluoride is a potent inhibitor of enolase. Mg^{2+} and inorganic phosphate are believed to form an ionic complex with fluoride ion, which is responsible for inhibition of the enzyme, apparently by interfering with the combination of the enzyme with its substrate (a Mg^{2+} plus 2-phosphoglycerate complex).

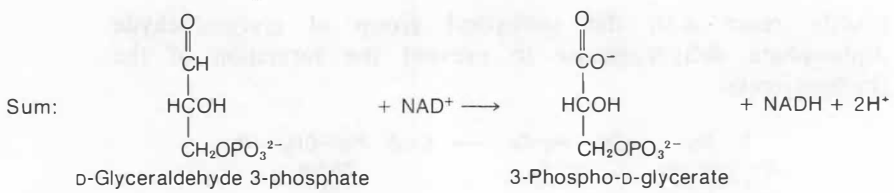
Arsenate is another important compound with respect to its effects on the glycolytic pathway. In the sense that it does not prevent flux through the steps of glycolysis, arsenate is not an inhibitor of the process. However, by causing arsenolysis at the step catalyzed by glyceraldehyde 3-phosphate dehydrogenase, arsenate prevents net synthesis of ATP by the glycolytic pathway. Arsenate looks a lot like inorganic phosphate, particularly to enzymes, and is able to substitute for inorganic phosphate in enzyme-catalyzed reactions. The result, in the case of the reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase, is the formation of a mixed anhydride of arsenic acid and the carboxyl group of 3-phosphoglycerate. This



compound is unstable, undergoing spontaneous hydrolysis to give 3-phosphoglycerate and inorganic arsenate. Hence, as is shown by



the sum of the last two reactions, glycolysis continues unabated in

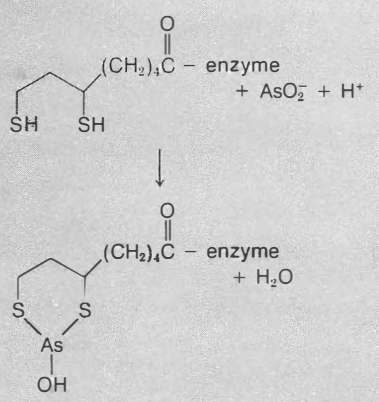


D-Glyceraldehyde 3-phosphate — 3-Phospho-D-glycerate

the presence of arsenate, but 1,3-bisphosphoglycerate is not formed, resulting in the loss of the capacity to synthesize ATP at the step catalyzed by 1,3-bisphosphoglycerate kinase. The consequence is that net ATP synthesis does not occur when glycolysis is carried out in the presence of arsenate, the ATP invested in the primary stage being only balanced by the ATP generated in the pyruvate kinase step. This means that in the presence of arsenate, glycolysis does not generate ATP, which can be used to meet the energy needs of a cell. This, along with the fact that arsenolysis also interferes with ATP formation by oxidative phosphorylation, makes arsenate a toxic compound (see Clin. Corr. 7.2).

CLIN. CORR. 7.2 ARSENIC POISONING

Most forms of arsenic are toxic, but the trivalent form, that is, arsenite or AsO_2^-, is much more toxic than the pentavalent form, that is, arsenate or $HAsO_4^{2-}$. Less ATP is produced whenever arsenate substitutes for inorganic phosphate in biological reactions. Arsenate competes for inorganic phosphate binding sites on enzymes, resulting in the formation of arsenate esters, which are unstable. Thus substrate level phosphorylation, for example, that occurring at the step glyceraldehyde 3-phosphate dehydrogenase, is prevented. Arsenite works by a completely different mechanism, involving the formation of a stable complex with enzyme bound lipoic acid:



7.4 REGULATION OF THE GLYCOLYTIC PATHWAY

Depending somewhat on the tissue under consideration, the regulatory enzymes of the glycolytic pathway are commonly considered to be hexokinase, phosphofructokinase, and pyruvate kinase. A summary of the important regulatory features of these enzymes is presented in Figure 7.6. A regulatory enzyme is defined as an enzyme that is subject to control by either allosteric effectors or covalent modulation. Both mechanisms are used by cells to control the most important of the regulatory enzymes. A regulatory enzyme can often be identified by determining whether the concentrations of the substrates and products within a cell indicate that the reaction catalyzed by the enzyme is close to equilibrium. An enzyme that is not subject to regulation will catalyze a "near-equilibrium" reaction, whereas a regulatory enzyme will catalyze a "nonequilibrium reaction" under intracellular conditions. This makes sense because flux through the step catalyzed by a regulatory enzyme is restricted because of controls imposed upon that enzyme. Whether an enzyme-catalyzed reaction is near equilibrium or nonequilibrium can be determined by comparing the established equilibrium constant for the reaction with the mass–action ratio as it exists within a cell. The

equilibrium constant for the reaction A + B → C + D is defined as

$$Keq = \frac{[C][D]}{[A][B]}$$

where the brackets indicate the concentrations at equilibrium. The mass–action ratio is calculated in a similar manner, except that the steady-state (ss) concentrations of reactants and products within the cell are used in the equation:

$$\text{Mass–action ratio} = \frac{[C]_{ss}[D]_{ss}}{[A]_{ss}[B]_{ss}}$$

If the mass–action ratio is approximately equal to the K_{eq}, the enzyme is said to be active enough to catalyze a near-equilibrium reaction and the enzyme is not considered subject to regulation. When the mass–action ratio is considerably different from the K_{eq}, the enzyme is said to catalyze a nonequilibrium reaction, is considered a

For the most part arsenic poisoning is explained by inhibition of those enzymes which require lipoic acid as a coenzyme. These include pyruvate dehydrogenase, α-ketoglutarate dehydrogenase, branched-chain α-keto acid dehydrogenase, and the glycine cleavage enzyme. Chronic arsenic poisoning, for example, accidentally from well water contaminated with arsenical pesticides or intentionally through the efforts of a murderer, is best diagnosed by determining the concentration of arsenic in the hair or fingernails of the victim. About 0.5 mg of arsenic would be found in a kilogram of hair from a normal individual. The hair of a person chronically exposed to arsenic could have 100 times as much.

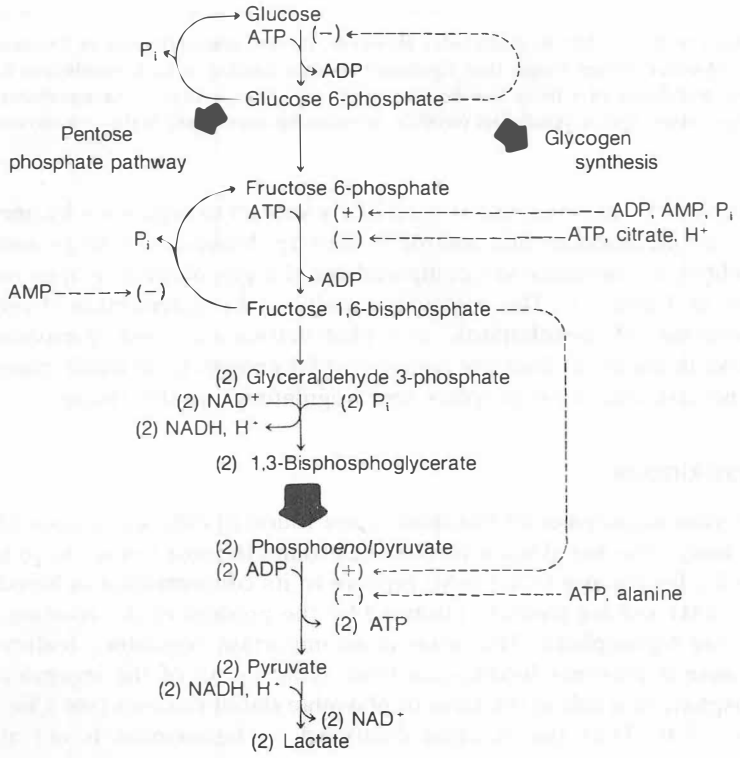


Figure 7.6
Important regulatory features of the glycolytic pathway.
Because of differences in isoenzyme distribution, not all tissues of the body have all of the regulatory mechanisms shown here.

Table 7.1 Apparent Equilibrium Constants and Mass–Action Ratios for the Reactions of Glycolysis and Gluconeogenesis in Liver

Reaction Catalyzed by	*Reaction in the Pathway of*		*Apparent Equilibrium Constant* (K'_{eq})	*Mass–Action Ratios*	*Considered Near-Equilibrium Reaction?*
	Glycolysis	*Gluconeogenesis*			
Glucokinase	Yes	No	2×10^3	0.02	No
Glucose 6-phosphatase	No	Yes	850 M	120 M	No
Phosphoglucoisomerase	Yes	Yes	0.36	0.31	Yes
6-Phosphofructokinase	Yes	No	1×10^3	0.09	No
Fructose bisphosphatase	No	Yes	530 M	19 M	No
Aldolase	Yes	Yes	13×10^{-5} M	12×10^{-7} M	Yes[a]
Glyceraldehyde 3-phosphate dehydrogenase + phosphoglycerate kinase	Yes	Yes	2×10^3 M^{-1}	0.6×10^3 M^{-1}	Yes
Phosphoglycerate mutase	Yes	Yes	0.1	0.1	Yes
Enolase	Yes	Yes	3.0	2.9	Yes
Pyruvate kinase	Yes	No	2×10^4	0.7	No
Pyruvate carboxylase + phospho*enol*pyruvate carboxykinase	No	Yes	7.0 M	1×10^{-3} M	No

[a] Reaction catalyzed by aldolase appears to be out of equilibrium by two orders of magnitude. However, in vivo concentrations of fructose 1,6-bisphosphate and glyceraldehyde 3-phosphate are so low (μM concentration range) that significant enzyme binding of both metabolites is believed to occur. Although only the total concentration of any metabolite of a tissue can be measured, only that portion of the metabolite that is not bound should be used in the calculations of mass–action ratios. This is usually not possible, introducing uncertainty in the comparison of in vitro equilibrium constants to in vivo mass–action ratios.

CLIN. CORR. 7.3 FRUCTOSE INTOLERANCE

Patients with hereditary fructose intolerance are deficient in the liver aldolase responsible for splitting fructose 1-phosphate into dihydroxyacetone phosphate and glyceraldehyde. Consumption of fructose by these patients results in the accumulation of fructose 1-phosphate and depletion of inorganic phosphate and ATP in the liver. The reactions involved are those catalyzed by fructokinase and the enzymes of oxidative phosphorylation:

"regulatory" enzyme, and is most likely subject to regulation by one or more mechanisms that control its activity. Mass–action ratios and equilibrium constants are compared for the glycolytic enzymes of liver in Table 7.1. The reactions catalyzed by glucokinase (liver isoenzyme of hexokinase), phosphofructokinase, and pyruvate kinase in the intact liver are considered far enough from equilibrium to indicate that these enzymes are "regulatory" in this tissue.

Hexokinase

Different isoenzymes of hexokinase are found in different tissues of the body. The hexokinase isoenzymes found in most tissues have a low K_m for glucose (<0.1 mM) relative to its concentration in blood (~5 mM) and are strongly inhibited by the product of the reaction, glucose 6-phosphate. The latter is an important regulatory feature because it prevents hexokinase from tying up all of the inorganic phosphate of a cell in the form of phosphorylated hexoses (see Clin. Corr. 7.3). Thus the reaction catalyzed by hexokinase is not at

equilibrium within cells that contain this enzyme because of the inhibition imposed by glucose 6-phosphate. Liver parenchymal cells are unique in that they contain glucokinase, an isoenzyme of hexokinase with strikingly different kinetic properties from the other hexokinases. This isoenzyme catalyzes the same reaction, that is, an ATP-dependent phosphorylation of glucose, but has a much higher K_m for glucose and is not subject to product inhibition by glucose 6-phosphate. The high K_m of glucokinase for glucose contributes to the capacity of the liver to "buffer" blood glucose levels. Glucose equilibrates readily across the plasma membrane of the liver, the concentration within the liver reflecting that of the blood. Since the K_m of glucokinase for glucose (~10 mM) is considerably greater than normal blood glucose concentrations (~5 mM), any increase in glucose concentration leads to a proportional increase in the rate of glucose phosphorylation by glucokinase (see Figure 7.7). Likewise, any decrease in glucose concentration leads to a proportional decrease in the rate of glucose phosphorylation. The result is that the liver uses glucose at a significant rate only when blood glucose levels are elevated. This buffering effect of liver glucokinase on blood glucose levels would not occur if glucokinase had the low K_m for glucose characteristic of other hexokinases and was, therefore, completely saturated at physiological concentrations of glucose (see Figure 7.7). On the other hand, a low K_m form of hexokinase is a good choice for tissues such as the brain in that it allows phosphorylation of glucose even when blood and tissue glucose concentrations are dangerously low.

In spite of the fact that glucokinase is not subject to inhibition by glucose 6-phosphate, the reaction catalyzed by glucokinase is not near-equilibrium under the intracellular conditions of liver cells (Ta-

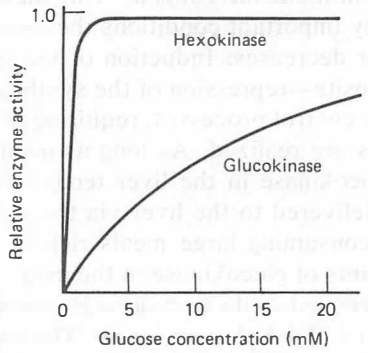


Figure 7.7
Comparison of the substrate saturation curves for hexokinase and glucokinase.

Fructose + ATP ⟶
fructose 1-phosphate + ADP
ADP + P_i + "energy provided
by electron transfer chain" ⟶ ATP
Net: P_i + fructose ⟶
fructose 1-phosphate

Tying up inorganic phosphate in the form of fructose 1-phosphate makes it impossible for liver mitochondria to generate ATP by oxidative phosphorylation. ATP levels fall precipitously, making it also impossible for the liver to carry out its normal work functions. Damage results to the cells in large part because they are unable to maintain normal ion gradients by means of the ATP-dependent cation pumps. The cells swell and eventually lose their internal contents by osmotic lysis.

Although patients with fructose intolerance are particularly sensitive to fructose, humans in general have a limited capacity to handle this sugar. The rate-limiting enzyme for fructose catabolism is the one defective in hereditary fructose intolerance, the fructose 1-phosphate aldolase. The capacity of the liver to phosphorylate fructose greatly exceeds its capacity to split fructose 1-phosphate. This means that fructose utilization by the liver is poorly controlled, that is, fructokinase should be subject to a regulatory mechanism, which would prevent it from depleting the liver of inorganic phosphate and ATP. Fructose was actually tried briefly in hospitals as a substitute for glucose with patients being maintained by parenteral nutrition. The rationale was that fructose would be a better source of calories than glucose because fructose utilization is relatively independent of the insulin status of a patient. Delivery of large amounts of fructose by intravenous feeding was soon found to result in severe liver damage. Similar attempts have been made to sub-

stitute sorbitol and xylitol for glucose. These sugars also tend to deplete the liver of ATP and, like fructose, should be used for parenteral nutrition with caution and only under special circumstances, if at all.

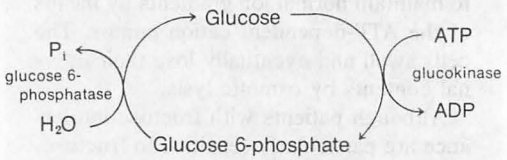


Figure 7.8
The phosphorylation of glucose followed by dephosphorylation constitutes a futile cycle in parenchymal cells of the liver.

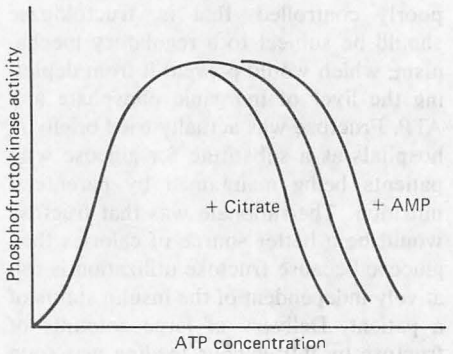


Figure 7.9
Effect of ATP, citrate, and AMP on the activity of phosphofructokinase.

ble 7.1). Part of the explanation lies in the rate restriction imposed by the high K_m of glucokinase for glucose. Another important factor is that the activity of glucokinase is opposed in liver by the enzyme glucose 6-phosphatase. Like glucokinase, this enzyme has an unusually high K_m (2.0 mM) with respect to the normal intracellular concentration (about 0.2 mM) of its primary substrate, glucose 6-phosphate. The result is that flux through this enzyme-catalyzed step is almost directly proportional to the intracellular concentration of glucose 6-phosphate. As shown in Figure 7.8, the combined action of glucokinase and glucose 6-phosphatase constitutes a futile cycle, that is, the sum of their reactions is simply the hydrolysis of ATP to give ADP and P_i without the performance of any work. It turns out that when blood glucose concentrations are about 5 mM, the activity of glucokinase is almost exactly balanced by the opposing activity of glucose 6-phosphatase. The result is that no net flux occurs in either direction. This futile cycling between glucose and glucose 6-phosphate is wasteful of ATP but, combined with the process of gluconeogenesis, contributes significantly to the "buffering" action of the liver on blood glucose levels. Furthermore, it provides a mechanism for preventing glucokinase from tying up all of the inorganic phosphate of the liver (see Clin. Corr. 7.3). An accumulation of glucose 6-phosphate in the liver does not inhibit glucokinase activity because glucose 6-phosphate is not an effective inhibitor of this isoenzyme of hexokinase. The effectiveness of glucokinase is nullified, however, because an increase in glucose 6-phosphate concentration increases the rate of glucose 6-phosphate hydrolysis by glucose 6-phosphatase. This helps prevent glucose 6-phosphate from accumulating faster than it can be used by other metabolic processes of the liver.

Glucokinase is an inducible enzyme. This means that under certain physiologically important conditions the *amount* of the enzyme either increases or decreases. Induction of the synthesis of an enzyme and the opposite—repression of the synthesis of an enzyme—are relatively slow control processes, requiring several hours before significant changes are realized. As long as insulin is also present, the amount of glucokinase in the liver tends to reflect how much glucose is being delivered to the liver via the portal vein. In other words, a person consuming large meals rich in carbohydrate will have greater amounts of glucokinase in the liver. The liver in which glucokinase has been induced can make a greater contribution to the lowering of elevated blood glucose levels. The exact details of how induction of glucokinase takes place at the gene level are not known, but it is clear that insulin is involved. The absence of insulin makes the liver of the diabetic patient deficient in glucokinase, in spite of high blood glucose levels, and this is one of the reasons why the liver

of the diabetic has less blood glucose "buffering" action (see Clin. Corr. 7.4).

Phosphofructokinase

Phosphofructokinase is the next regulatory enzyme of the glycolytic pathway. Much evidence suggests that this is the rate-limiting enzyme of glycolysis in most tissues. It is subject to regulation by a number of allosteric effectors, including ATP, which is an inhibitor of the enzyme; citrate, which potentiates ATP inhibition; and AMP, which relieves ATP inhibition. The inhibitory effect of ATP (shown in Figure 7.9) is somewhat unusual, since it is a primary substrate for the reaction catalyzed by the enzyme. ATP promotes the reaction at low concentrations, as would be expected for a substrate, but inhibits at high concentrations. The potentiating effect of citrate on ATP inhibition of phosphofructokinase, and the relief affected by AMP, are shown in Figure 7.9. Hydrogen ions (i.e., low pH) also inhibit phosphofructokinase. All of these effectors of phosphofructokinase are believed to be physiologically important, strong evidence for this conclusion coming from application of the crossover theorem to the glycolytic pathway.

For the hypothetical pathway of $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \ldots$, the crossover theorem proposes that an inhibitor that partially inhibits the conversion of C to D will cause a "crossover" in the metabolite profile between C and D. This means that when the steady-state concentrations of the intermediates in the presence and absence of an inhibitor are compared, the concentrations of the intermediates before the site of inhibition should increase in response to the inhibitor, whereas those after the site should decrease. Crossover plots are constructed by setting the concentrations of all intermediates without some effector of the pathway equal to 100%. The concentrations of the intermediates observed in the presence of the effector are then expressed as percentages of these values. Expected results with positive and negative effectors are shown in Figure 7.10. The effect of returning the perfused rat heart from an anoxic condition to a well-oxygenated state is also shown. This transition with the perfused rat heart is known to establish new steady-state concentrations of the glycolytic intermediates, the flux being much greater through the glycolytic pathway in the absence of oxygen. Under experimental conditions used for Figure 7.10*c*, the perfused hearts consumed glucose at rates some 20 times greater in the absence than in the presence of oxygen. This example illustrates what is known as the Pasteur effect, defined as the inhibition of glucose utilization and lactate accumulation by the initiation of respiration (oxygen consumption). This is readily understandable on a

CLIN. CORR. 7.4 DIABETES

Diabetes mellitus is a chronic disease characterized by derangements in carbohydrate, fat, and protein metabolism. Two types are recognized clinically—the juvenile-onset or insulin-dependent type (see Clin. Corr. 14.7) and the maturity-onset or insulin-independent type (see Clin. Corr. 14.9).

The oral glucose tolerance test is commonly used for the diagnosis of diabetes. It consists of determining the blood glucose level in the fasting state and at intervals of 30–60 min for 2 h or more after consuming a 100 g carbohydrate meal. In a normal individual blood glucose returns to normal levels within 2 h after ingestion of the carbohydrate meal. In the diabetic blood glucose will reach a higher level and remain elevated for longer periods of time, depending upon the severity of the disease. However, many factors may contribute to an abnormal glucose tolerance test. The patient must have consumed a high carbohydrate diet for the preceding 3 days, presumably to allow for induction of glucose-utilizing pathways and enzymes, for example, glucokinase, fatty acid synthetase, acetyl CoA carboxylase. In addition, almost any infection (colds, etc.) and less well-defined "stress," possibly by effects on the sympathetic nervous system, can result in (transient) abnormalities of the glucose tolerance test. In fact, a majority of people screened for diabetes with a glucose tolerance test who have an abnormal result will not develop diabetes. Because of the cited problems with the glucose tolerance test, elevation of the fasting glucose level should probably be the sine qua non for the diagnosis of diabe-

tes. Glucose uptake by cells of insulin-sensitive tissues, that is, muscle and adipose, is decreased in the diabetic state. Insulin is required for glucose uptake by these tissues, and the diabetic either lacks insulin or has developed "insulin resistance" in these tissues. Resistance to insulin is an abnormality of the insulin receptor or in subsequent steps mediating the metabolic effects of insulin. Parenchymal cells of the liver do not require insulin for glucose uptake. Without insulin, however, the liver has diminished enzymatic capacity to remove glucose from the blood. This is explained in large part by decreased glucokinase activity, along with other key enzymes of glycogenesis and the glycolytic pathway.

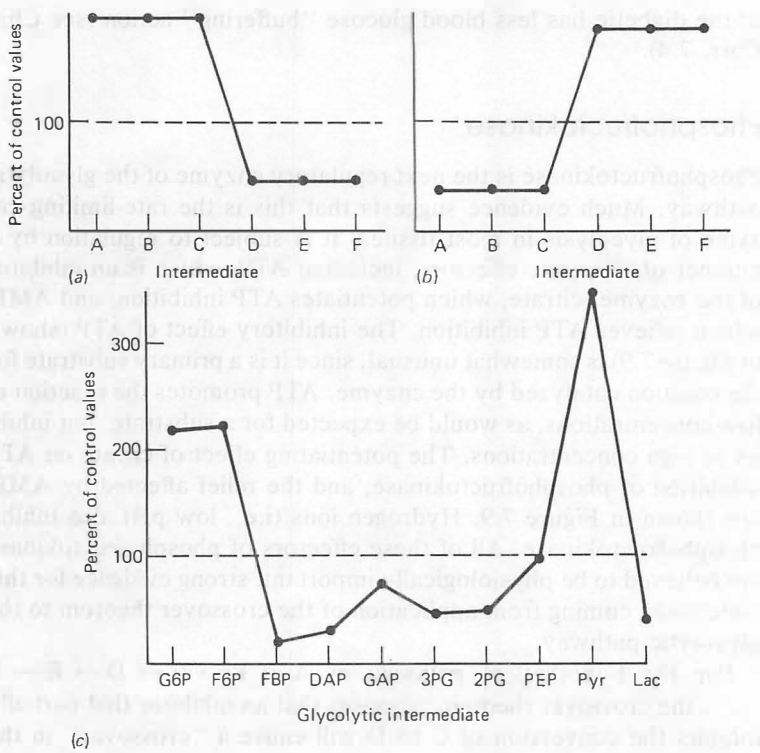


Figure 7.10

Crossover analysis is used to locate sites of regulation of a metabolic pathway.

(a) Theoretical effect of an inhibitor of the C to D step in the pathway of $A \rightarrow B \rightleftharpoons C \rightarrow D \rightleftharpoons E \rightarrow F$. *Steady-state concentrations of all intermediates of the pathway without the inhibitor present are arbitrarily set equal to 100%. Steady-state concentrations of all intermediates when the inhibitor is present are then expressed as percentages of the control values. (b) Theoretical effect of an activator of the C to D step in the pathway of* $A \rightarrow B \rightleftharpoons C \rightarrow D \rightleftharpoons E \rightarrow F$. *(c) Effect of oxygen on the relative steady-state concentrations of the intermediates of the glycolytic pathway in the perfused rat heart. The changes in concentrations of metabolites of hearts perfused without oxygen caused by subsequent perfusion with oxygen (95%* O_2, *5%* CO_2*) are recorded as percentages of the anoxic values. Oxygen strongly inhibits glucose utilization and lactate production under such conditions. The dramatic increase in pyruvate concentration occurs as a consequence of greatly increased utilization of cytosolic NADH by the shuttle systems. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; DAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; and Lac, lactate.*

From J. R. Williamson, *J. Biol. Chem.*, 241, 5026 (1966).

thermodynamic basis, the complete oxidation of glucose to CO_2 and H_2O yielding so much more ATP than anaerobic glycolysis:

Glycolysis: D-Glucose + $2ADP^{3-}$ + $2P_i^{2-}$ ⟶ 2 L-lactate$^-$ + $2ATP^{4-}$

Complete Oxidation: D-Glucose + $6O_2$ + $38ADP^{3-}$ + $38P_i^{2-}$ + $38H^+$ ⟶ $6CO_2$ + $6H_2O$ + $38ATP^{4-}$

ATP is used by a cell only to meet its metabolic demand, that is, to provide the necessary energy for the work processes (metabolic demand) inherent to that cell. Since so much more ATP is produced from glucose in the presence of oxygen, much less glucose has to be consumed to meet the metabolic demand of the cell. The "crossover" at the conversion of fructose 6-phosphate to fructose 1,6-bisphosphate argues that oxygen imposes an inhibition at the level of phosphofructokinase. This can be readily rationalized on the basis that ATP is a well-recognized inhibitor of phosphofructokinase, and more ATP can be generated in the presence of oxygen than in the absence. Experimentally, however, things are not quite that simple. ATP levels have not been found to change greatly between these two conditions (in the experiment of Figure 7.10*c*, ATP increased from 4.7 μmol/g wet weight in the absence of oxygen to 5.6 μmol/g wet weight in the presence of oxygen). Such a small difference in ATP concentration cannot account for the large change in flux through phosphofructokinase. However, much greater changes, percentagewise, occur in the concentrations of AMP and P_i, effectors that relieve ATP inhibition of phosphofructokinase. The changes that occur in the steady-state concentrations of AMP and P_i when oxygen is introduced into the systems are exactly what might have been predicted, that is, the levels of both go down dramatically from 60 to 80%. These changes increase the extent of ATP inhibition of phosphofructokinase, greatly suppress glycolytic activity, and account in large part for the Pasteur effect. AMP levels automatically go down in a cell when ATP levels increase. Although this is not intuitively obvious, the reason is simple enough. The sum of the adenine nucleotides in a cell, that is, ATP + ADP + AMP, is nearly constant under most physiological conditions, but the relative concentrations are such that the ATP concentration is always much greater than the AMP concentration. Furthermore, the adenine nucleotides are maintained in equilibrium in the cytosol through the action of an enzyme called nucleoside monophosphokinase, which catalyzes the following reaction: $2ADP \rightleftharpoons ATP + AMP$. The equilibrium constant (K'_{eq}) for this reaction is given by the expression:

$$K'_{eq} = \frac{[ATP][AMP]}{[ADP]^2}$$

Since this reaction is "near-equilibrium" under intracellular conditions, the concentration of AMP is given by

$$[AMP] = \frac{K'_{eq}[ADP]^2}{[ATP]}$$

Due to the fact that intracellular [ATP] $\gg$ [ADP] $\gg$ [AMP], a small decrease in [ATP] causes a substantially greater increase in [ADP]; and, since [AMP] is related to the square of the [ADP], an even greater increase in [AMP]. Because of this relationship, a small decrease in ATP concentration is amplified into a much larger (percentage) change in AMP concentration. This makes AMP an excellent "sensor" of the energy status of the cell and allows it to function as an important allosteric effector of phosphofructokinase activity. Furthermore, AMP influences in yet another way the effectiveness of the reaction catalyzed by phosphofructokinase. An enzyme called fructose bisphosphatase catalyzes an irreversible reaction, which opposes that of phosphofructokinase:

$$\text{Fructose 1,6-bisphosphate} + H_2O \longrightarrow \text{fructose 6-phosphate} + P_i$$

This enzyme sits "cheek by jowl" with phosphofructokinase in the cytosol of many cells. Together they catalyze a futile cycle (ATP $\rightarrow$ ADP + P_i + "heat"), and, at the very least, they decrease the "effectiveness" of one another. The AMP concentration is a perfect sensor of the energy status of the cell—not only because AMP activates phosphofructokinase but because AMP also *inhibits* fructose bisphosphatase. The result is that a small decrease in ATP concentration triggers, via the increase in AMP concentration, a large increase in the net conversion of fructose 6-phosphate into fructose 1,6-bisphosphate. This increases the glycolytic flux by increasing the amount of substrate available for the splitting stage. In cells containing hexokinase, it also results in greater phosphorylation of glucose because a decrease in fructose 6-phosphate automatically causes a decrease in glucose 6-phosphate, which, in turn, results in less inhibition of hexokinase activity.

The decrease in lactate production in response to the onset of respiration is another feature of the Pasteur effect that can be readily explained. The most important factor is the decreased glycolytic flux caused by oxygen; however, secondary factors include competition between lactate dehydrogenase and the mitochondrial pyruvate dehydrogenase for pyruvate, as well as competition between lactate

dehydrogenase and the shuttle systems for NADH. For the most part, lactate dehydrogenase loses the competition in the presence of oxygen.

It would be natural to suspect that lactate, as the end product of glycolysis, would inhibit the rate-limiting enzyme of the glycolytic pathway. It does not. However, protons, the other glycolytic end product, do inhibit phosphofructokinase. As shown in Figure 7.11, glycolysis in effect generates lactic acid and the cell must dispose of it as such. This accounts for why excessive glycolysis in the body lowers blood pH and leads to an emergency medical situation termed lactic acidosis (see Clin. Corr. 7.5). Plasma membranes of cells appear to contain either an antiport for lactate and hydroxide ions or a symport for lactate and protons (experimentally they are difficult to distinguish). Regardless of the exact mechanism, lactic acid is released from the cell into the bloodstream. This ability to transport lactic acid out of the cell is a defense mechanism, preventing the pH from getting so low that everything becomes pickled (see Clin. Corr. 7.6). The sensitivity of phosphofructokinase to protons is also part of this mechanism. Protons are able to shut off glycolysis, the process responsible for decreasing the pH. Note that transport of lactic acid out of a cell requires that blood be available to the cell in order to carry this "end product" away. When blood flow to a group of cells is inadequate, for example, in heavy exercise of a skeletal muscle or an attack of angina pectoris in the case of the heart, protons cannot escape from the cells fast enough. Yet the need for ATP within such cells, because of the lack of oxygen, may partially override the inhibition of phosphofructokinase by protons. The unabated accumulation of protons then results in pain, which, in the case of skeletal muscle, can be relieved by simply terminating the exercise. In the case of the heart, rest or pharmacologic agents that increase blood flow or decrease the need for ATP within the myocytes may be effective (see Clin. Corr. 7.7).

Many tissues prefer to use fatty acids and ketone bodies as oxidizable fuels in place of glucose. Most of these tissues have the capacity to use glucose but actually prefer to oxidize fatty acids and ketone bodies. This unselfish act helps preserve glucose for those tissues, such as brain, that are absolutely dependent upon sugar as an energy source. The mechanism responsible for this preference is relatively simple. Oxidation of both fatty acids and ketone bodies elevates the levels of cytosolic citrate, which inhibits phosphofructokinase. The result is decreased glucose utilization by the tissue when fatty acids or ketone bodies are available.

There is apparently yet another way to regulate phosphofructokinase, the details of which are only now beginning to be unraveled. Phosphofructokinase appears to be an interconvertible enzyme, its

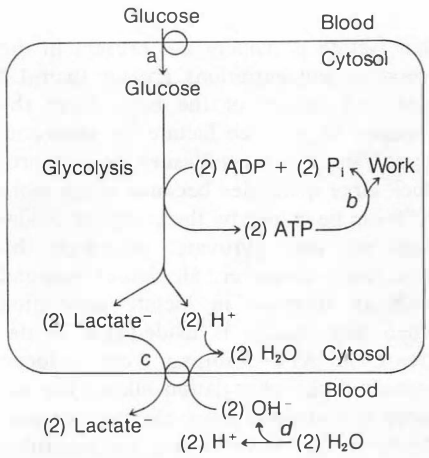


Figure 7.11
Unless lactate formed by glycolysis is released from the cell, the intracellular pH is decreased as a consequence of the accumulation of intracellular lactic acid.
The low pH decreases phosphofructokinase activity so that further lactic acid production by glycolysis is shut off. a, glucose transport into the cell; b, all work performances which convert ATP back to ADP and P_i; c, lactate–hydroxide ion antiport; d, ionization of water in the blood to give hydroxide ions for exchange with lactate.

CLIN. CORR. 7.5 LACTIC ACIDOSIS

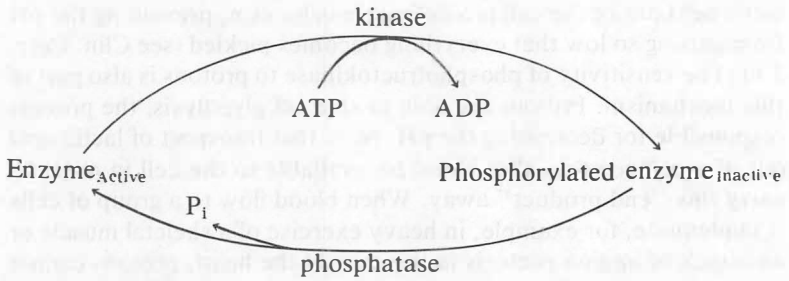
This problem is characterized by elevated blood lactate levels, usually greater than 5 mM, along with decreased blood pH and bicarbonate concentrations. Lactic acidosis is the most commonly encountered form of metabolic acidosis and can be the consequence of overproduction of lactate, underutilization of lactate, or both. Lactate production is normally balanced by lactate utilization, with the result

that lactate is usually not present in the blood at concentrations greater than 1.2 mM. All tissues of the body have the capacity to produce lactate by anaerobic glycolysis, but most tissues do not produce large quantities because much more ATP can be gained by the complete oxidation of the pyruvate produced by glycolysis. However, all tissues respond with an increase in lactate generation when oxygenation is inadequate. A decrease in ATP resulting from reduced oxidative phosphorylation allows the activity of phosphofructokinase to increase. These tissues have to rely on anaerobic glycolysis for ATP production under such conditions and this results in lactic acid production. A good example is muscle exercise, which can deplete the tissue of oxygen and cause an overproduction of lactic acid. Tissue hypoxia occurs, however, in all forms of shock, during convulsions, and in diseases involving circulatory and pulmonary failure.

The major fate of lactate in the body is either complete combustion to CO_2 and H_2O or conversion back to glucose by the process of gluconeogenesis. Both require oxygen. Decreased oxygen availability, therefore, increases lactate production and decreases lactate utilization. The latter can also be decreased by liver diseases, ethanol, and a number of other drugs. Phenformin, a drug which was once used to treat the hyperglycemia of insulin-independent diabetes, was well-documented to induce lactic acidosis in certain patients.

Bicarbonate is usually administered in an attempt to control the acidosis associated with lactic acid accumulation. The key to successful treatment, however, is to find and eliminate the cause of the overproduction and/or underutilization of lactic acid and most often involves the restoration of circulation of oxygenated blood.

activity subject to regulation by phosphorylation and dephosphorylation. Although research on this subject is in a state of flux at the moment, it appears that a kinase exists that renders phosphofructokinase inactive by phosphorylation of specific sites on the enzyme. By analogy to other interconvertible enzymes, a phosphoprotein phosphatase must exist that activates the enzyme by dephosphorylation (see scheme). This mechanism for regulating phosphofructokinase appears unique to the liver and probably explains why the hormone glucagon inhibits hepatic glycolysis and stimulates, as discussed below, hepatic gluconeogenesis. Presumably the kinase involved is directly or indirectly activated (or the phosphatase inhibited) by cAMP, the second messenger of glucagon action, but it is too early to be sure about that. However, as discussed in Chapter 14, this seems very likely to be the story.



That phosphofructokinase catalyzes the rate-limiting step of the glycolytic pathway may seem to be an enigma. Usually we think of the first step as the most logical choice for the rate-limiting step of a pathway. The thing to notice is that phosphofructokinase catalyzes the first *committed* step of the glycolytic pathway (see Figure 7.6). This step commits the cell to the metabolism of glucose by the glycolytic pathway. The reaction catalyzed by phosphoglucoisomerase is readily reversible and most cells can use glucose 6-phosphate in a number of ways, that is, for glycogen synthesis and/or in the pentose phosphate pathway (see Figure 7.6).

Pyruvate Kinase

Pyruvate kinase is another regulatory enzyme of glycolysis (see Clin. Corr. 7.8). However, as with hexokinase, the reaction catalyzed by pyruvate kinase has to be considered a secondary site of regulation of glycolysis. This enzyme is drastically inhibited by physiological concentrations of ATP, so much so that its potential activity is never fully realized under physiological conditions. The isoenzyme found in liver is greatly activated by fructose 1,6-

CLIN. CORR. 7.6 PICKLED PIGS AND MALIGNANT HYPERTHERMIA

In patients with malignant hyperthermia, a variety of agents, especially the widely used general anesthetic halothane, will produce a dramatic rise in body temperature, metabolic and respiratory acidosis, hyperkalemia, and muscle rigidity. This genetic abnormality occurs in only about 1 person in 50,000 to 100,000 and is transmitted as an autosomal dominant trait. Death often results the first time a susceptible person is anesthetized. Onset occurs within minutes of drug exposure and the hyperthermia must be recognized immediately. Packing the patient in ice is effective and should be accompanied by measures to combat acidosis. The drugs procainamide and dantrolene are also effective.

A phenomenon similar, if not identical, to malignant hyperthermia is known to occur in pigs. Pigs with this problem, called porcine stress syndrome, respond poorly to stress. This genetic disease usually manifests itself as the pig is being shipped to market. Pigs with the syndrome can be identified by exposure to halothane, which triggers the same response seen in patients with malignant hyperthermia. The meat of pigs that have died as a result of the syndrome is pale, watery, and of very low pH (i.e., nearly pickled).

Muscle is considered the site of the primary lesion in both malignant hyperthermia and porcine stress syndrome. In response to halothane the skeletal muscles become rigid and generate heat and lactic acid. Although much experimental work has been conducted, the biochemical basis for the increased heat production remains obscure. Heat produced by glycolytic activity and muscle contraction is not believed sufficient to explain the dramatic increase in body temperature. Uncontrolled futile cycling in which ATP hydrolysis is greatly accelerated has been suggested to be involved:

$$ATP + H_2O \rightarrow ADP + P_i + \text{heat}$$

Indeed halothane has been shown to accelerate futile cycling at the level of phosphofructokinase-fructose bisphosphatase in muscles of pigs with porcine stress syndrome. Perhaps one of these regulatory enzymes will be found defective with respect to allosteric effector control in patients with malignant hyperthermia. There is also evidence that the sarcoplasmic reticulum of such patients may be defective and that the anesthetic triggers inappropriate release of Ca^{2+} from the sarcoplasmic reticulum. This could result in uncontrolled stimulation of a number of heat producing processes, that is, myosin ATPase, glycogenolysis, glycolysis, and cyclic uptake and release of Ca^{2+} by mitochondria and sarcoplasmic reticulum.

CLIN. CORR. 7.7 ANGINA PECTORIS AND MYOCARDIAL INFARCTION

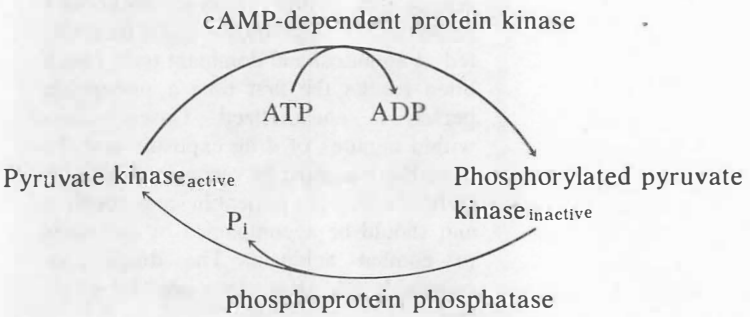
Chest pain correlating with reversible myocardial ischemia is termed angina pectoris. The pain is the result of an imbalance between demand for and supply of blood flow to cardiac muscles and is most commonly caused by coronary artery

obstructive disease. The patient experiences a heavy squeezing pressure or ache substernally, often radiating to either shoulder and arm or occasionally to the jaw or neck. Attacks occur with exertion, last from 1 to 15 min, and are relieved by rest. The coronary arteries involved are obstructed by atherosclerosis (i.e., lined with characteristic fatty deposits) or less commonly narrowed by spasm. Myocardial infarction occurs if the ischemia persists long enough to cause severe damage (necrosis) to the heart muscle. In myocardial infarction, tissue death occurs and the characteristic pain is longer lasting, and often more severe.

Nitroglycerin and other nitrates are frequently prescribed to relieve the pain caused by the myocardial ischemia of angina pectoris. These drugs can be used prophylactically, enabling patients to participate in activities that would otherwise precipitate an attack of angina. Nitroglycerin may work in part by causing dilation of the coronary arteries, improving oxygen delivery to the heart and washing out lactic acid. Probably more important is the effect of nitrates on the peripheral circulation. Nitrates relax smooth muscle, causing vasodilation throughout the body. This reduces arterial pressure and allows blood to accumulate in the veins. The result is decreased return of blood to the heart, which reduces the volume of the heart, which reduces the energy requirement of the heart. In addition, the heart empties itself against less pressure, which also spares energy The overall effect is a lowering of the oxygen requirement of the heart, bringing it in line with the oxygen supply via the diseased coronary arteries.

The coronary artery bypass operation is used in severe cases of angina that cannot be controlled by medication. In this operation veins are removed from the leg and interposed between the aorta and coronary arteries of the heart. The purpose is to

bisphosphate, thereby linking regulation of pyruvate kinase to what is happening at the level of phosphofructokinase. Thus, if conditions favor increased flux through phosphofructokinase, the level of fructose 1,6-bisphosphate increases and acts as a feed-forward activator of pyruvate kinase. The liver enzyme is also subject to covalent modulation, being active in the dephosphorylated state and inactive in the phosphorylated state. Inactivation of pyruvate kinase by phosphorylation is a function of cAMP-dependent protein kinase in the liver. Glucagon inhibition of hepatic glycolysis and stimulation of hepatic gluconeogenesis are explained in part by the elevation of cAMP levels caused by this hormone. This aspect is explored more thoroughly under the section of this chapter on gluconeogenesis, and in Chapter 14.



Pyruvate kinase, like glucokinase, is induced to higher steady-state concentrations in the liver by the combination of high carbohydrate intake and high insulin levels. This is a major reason why the liver of the well-fed individual has much greater capacity for utilizing carbohydrate than a fasting or diabetic person (see Clin. Corr. 7.4).

7.5 GLUCONEOGENESIS

The Importance of Glucose Synthesis

The net synthesis or formation of glucose from a large variety of substrates is termed gluconeogenesis. This includes the use of various amino acids, lactate, pyruvate, propionate, glycerol, galactose, and fructose as sources of carbon for the pathway (see Figure 7.12). Glycogenolysis, that is, the formation of glucose or glucose 6-phosphate from glycogen, should be carefully differentiated from

gluconeogenesis; glycogenolysis refers to

$$\text{Glycogen or } (\text{glucose})_n \dashrightarrow n \text{ molecules of glucose}$$

and thus does not correspond to de novo or net synthesis of glucose, the hallmark of the process of gluconeogenesis.

The capacity to synthesize glucose is crucial for the survival of humans and other animals. Blood glucose levels have to be maintained to support the metabolism of those tissues that use glucose as their primary substrate (see Clin. Corr. 7.9). This includes brain, red blood cells, kidney medulla, lens and cornea of the eye, testis, and a number of minor tissues. Gluconeogenesis enables the maintenance of blood glucose levels long after all dietary glucose in the form of glycogen and sucrose has been absorbed and completely oxidized. Indeed, gluconeogenesis is important to the human every day, making it possible for us to make it through the night and from meal to meal without nibbling on a source of carbohydrate continuously.

The Cori and Alanine Cycles

Two important cycles between tissues are recognized to involve the process of gluconeogenesis. The Cori cycle and the alanine cycle, given in Figure 7.13, consist of gluconeogenesis in the liver followed by glycolysis in a peripheral tissue. The purpose of both is to provide a mechanism for continuously supplying glucose to tissues that are dependent on it as their primary energy source. The cycles are only functional, however, between the liver and tissues that do not completely oxidize glucose to CO_2 and H_2O. In order to participate in these cycles, the peripheral tissue must release either alanine or lactate as the end product of glycolysis. The recycled three-carbon intermediate is the major difference between the Cori cycle and the alanine cycle, carbon returning to the liver in the form of lactate in

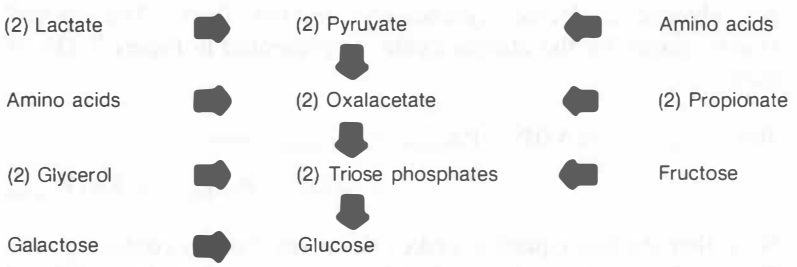


Figure 7.12
Abbreviated pathway of gluconeogenesis, illustrating the major substrate precursors for the process.

bypass the portion of the artery diseased by atherosclerosis and provide the affected tissue with a greater blood supply. Remarkable relief from angina can be achieved by this operation, with the patient being able to return to normal productive life in some cases.

CLIN. CORR. 7.8 PYRUVATE KINASE DEFICIENCY AND HEMOLYTIC ANEMIA

Mature erythrocytes are absolutely dependent upon glycolytic activity for ATP production. ATP is needed for the ion pumps, especially the Na^+, K^+–ATPase, which maintain the biconcave disk shape of erythrocytes, a characteristic that helps erythrocytes slip through the capillaries as they deliver oxygen to the tissues. Without ATP the cells swell and lyse. Anemia due to excessive erythrocyte destruction is referred to as hemolytic anemia. Pyruvate kinase deficiency is rare, but is by far the most common genetic defect of the glycolytic pathway known to cause hemolytic anemia. Although most pyruvate kinase-deficient patients have 5–25% of normal red blood cell pyruvate kinase levels, flux through the glycolytic pathway is restricted severely, resulting in markedly lower ATP concentrations. The expected crossover of the glycolytic intermediates is observed, that is, those intermediates proximal to the pyruvate kinase-catalyzed step accumulate, whereas pyruvate and lactate concentrations decrease. Normal ATP levels are observed in reticulocytes of patients with this disease. Although deficient in pyru-

vate kinase, these "immature" red blood cells have mitochondria and can generate ATP by oxidative phosphorylation. Maturation of reticulocytes into red blood cells results in the loss of mitochondria and complete dependence on glycolysis for ATP production. Since glycolysis is defective, the mature cells are lost rapidly from the circulation. Anemia results because the cells cannot be replaced rapidly enough by erythropoiesis.

CLIN. CORR. 7.9 HYPOGLYCEMIA AND PREMATURE INFANTS

Premature and small-for-gestational-age neonates have a greater susceptibility to hypoglycemia than full-term, appropriate-for-gestational-age infants. Several factors appear to be involved. Children in general are more susceptible than adults to hypoglycemia, simply because they have larger brain to body weight ratios and the brain utilizes disproportionately greater amounts of glucose than the rest of the body. Newborn infants have a limited capacity for ketogenesis, apparently because the transport of long-chain fatty acids into liver mitochondria of the neonate is poorly developed. Since ketone body use by the brain is directly proportional to the circulating ketone body concentration, the neonate is unable to spare glucose to any significant extent by using ketone bodies. The consequence is that the neonate's brain is almost completely dependent upon glucose obtained from liver glycogenolysis and gluconeogenesis.

The capacity for hepatic glucose synthesis from lactate and alanine is also limited.

the Cori cycle but in the form of alanine in the alanine cycle. Another difference is that the NADH generated by glycolysis in the alanine cycle cannot be used to reduce pyruvate to lactate. In tissues that have mitochondria, the electrons of NADH can be transported into the mitochondria by the malate–aspartate shuttle or the glycerol phosphate shuttle for the synthesis of ATP by oxidative phosphorylation:

$$NADH + H^+ + \tfrac{1}{2}O_2 + 3ADP + 3P_i \longrightarrow NAD^+ + 3ATP$$

or

$$FADH_2 + \tfrac{1}{2}O_2 + 2ADP + 2P_i \longrightarrow FAD + 2ATP$$

The consequence is that six to eight molecules of ATP can be formed per glucose in peripheral tissues that participate in the alanine cycle. This stands in contrast to the Cori cycle, in which only two ATPs per glucose are produced. Inspection of Figure 7.13*a* will reveal that the overall stoichiometry for the Cori cycle is

$$6ATP_{liver} + 2(ADP + P_i)_{red\ blood\ cells} \longrightarrow 6(ADP + P_i)_{liver} + 2ATP_{red\ blood\ cells}$$

The six ATPs are needed in the liver to provide the energy necessary for glucose synthesis. The alanine cycle also in effect transfers high energy phosphate from the liver to the peripheral tissues and, because of the 6 to 8 ATPs produced per glucose, is an energetically more efficient cycle. However, as shown in Figure 7.13*b*, the participation of alanine in the cycle presents the liver with amino nitrogen, which must be disposed of as urea. In terms of ATP, urea synthesis is expensive (4 ATP molecules per urea molecule). The concurrent need for urea synthesis results in more ATP being needed per glucose molecule synthesized in the liver. The overall stoichiometry for the alanine cycle, as presented in Figure 7.13*b*, is then

$$10ATP_{liver} + 6\text{–}8(ADP + P_i)_{muscle} + O_{2\ muscle} \longrightarrow 10(ADP + P_i)_{liver} + 6\text{–}8ATP_{muscle}$$

Note that the last equation makes the point that, in contrast to the Cori cycle, oxygen and mitochondria are required in the peripheral tissue for participation in the alanine cycle.

Liver was used as the example in Figure 7.13 because it is the most important gluconeogenic tissue. The kidneys, on a gram wet

weight basis, have about the same capacity for the process. However, the liver is the largest organ in the body, exceeding the combined weight of the kidneys by a factor of 4, and thus contributes much more to the maintenance of blood glucose levels by gluconeogenesis. Certain muscle fibers may have the capacity for limited gluconeogenesis. Since the adult human has 18 times more

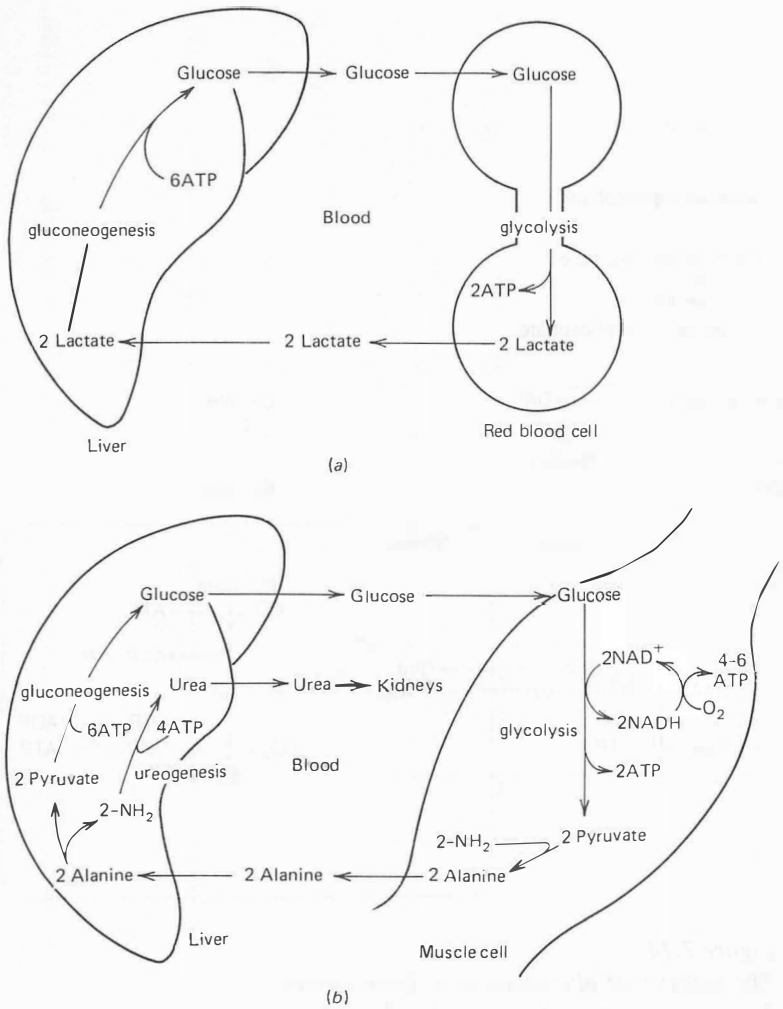


Figure 7.13
Relationship between gluconeogenesis in the liver and glycolysis in the rest of the body.
(*a*) *Cori cycle.* (*b*) *Alanine cycle.*

ited in newborn infants. This is because the rate-limiting enzyme phospho*enol*-pyruvate carboxykinase is present in very low amounts during the first few hours after birth. Induction of this enzyme to the level required to prevent hypoglycemia during the stress of fasting requires several hours. Premature and small-for-gestational-age infants are believed to be more susceptible to hypoglycemia than normal infants because of smaller stores of liver glycogen. Fasting depletes their glycogen stores more rapidly, making these neonates more dependent on gluconeogenesis than normal infants.

muscle mass than liver, glucose synthesis in muscle may eventually be shown to be quantitatively important. However, muscle tissue lacks glucose 6-phosphatase, the enzyme that catalyzes the last step of the gluconeogenic pathway. Thus, any gluconeogenesis occurring in muscle should be pictured as taking place in order to help replenish glycogen stores in this tissue, rather than for the production of free glucose for the maintenance of blood sugar levels.

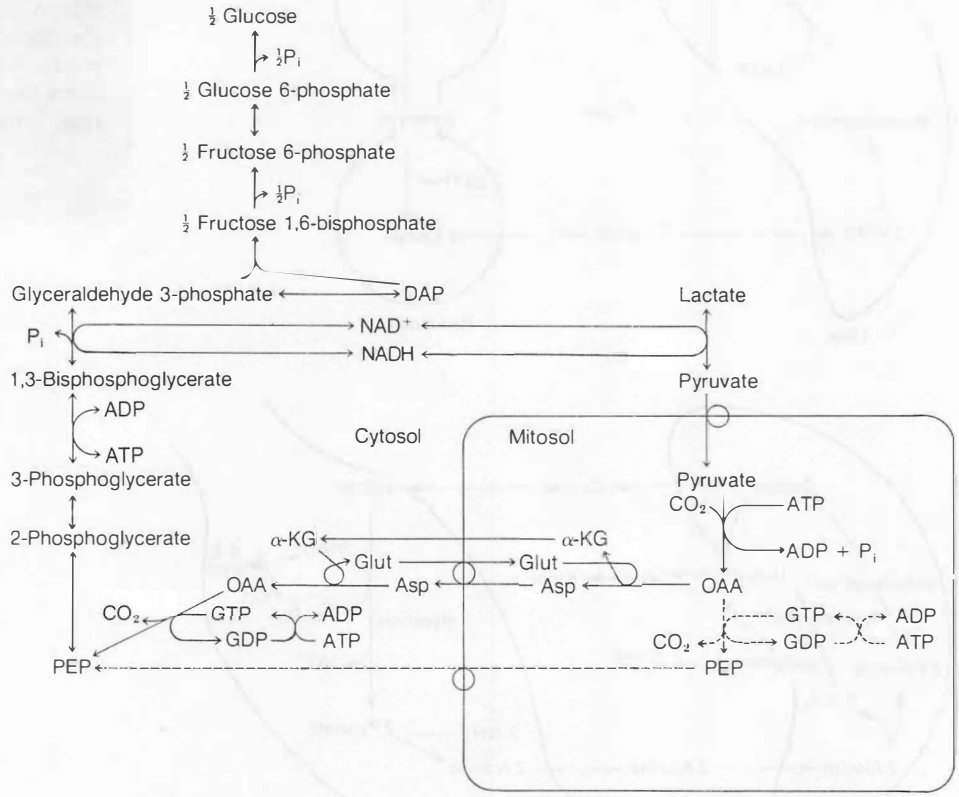


Figure 7.14
The pathway of gluconeogenesis from lactate.
*The involvement of the mitochondrion in the process is indicated in the figure. Dashed arrows refer to an alternate route which employs mitosolic phospho*enol*pyruvate carboxykinase rather than the cytosolic isoenzyme. Abbreviations: OAA, oxalacetate; α-KG α-ketoglutarate; PEP, phospho*enol*pyruvate; and DAP, dihydroxy acetone phosphate.*

Pathway Responsible for Glucose Synthesis

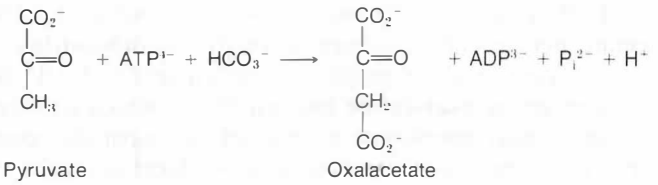
Lactate gluconeogenesis is an ATP-requiring process with the overall equation of

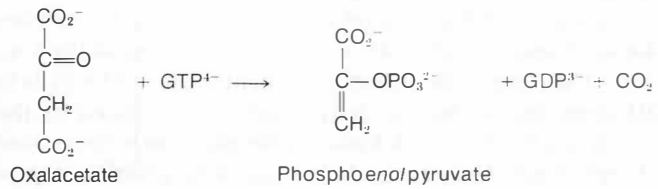
$$2\ \text{L-Lactate}^- + 6\text{ATP}^{4-} \longrightarrow \text{glucose} + 6\text{ADP}^{3-} + 6\text{P}_i^{2-} + 4\text{H}^+$$

Many of the enzymes of the glycolytic pathway are common to the gluconeogenic pathway but, it is obvious from the overall equation for glycolysis,

$$\text{Glucose} + 2\text{ADP}^{3-} + 2\text{P}_i^{2-} \longrightarrow 2\ \text{L-lactate}^- + 2\text{ATP}^{4-}$$

that additional reactions have to be involved. Also, as pointed out in the discussion of glycolysis, certain steps of the glycolytic pathway are irreversible under intracellular conditions. For glucose synthesis to occur, the irreversible steps of the glycolytic pathway are replaced by irreversible steps of the gluconeogenic pathway. The reactions of the gluconeogenic pathway for glucose synthesis from lactate are given in Figure 7.14. The initial step is the conversion of lactate to pyruvate by lactate dehydrogenase. NADH is generated, but its conversion back to NAD^+ is no problem for the process because NADH is needed for a subsequent step of the pathway. Pyruvate cannot be converted into phospho*enol*pyruvate by reversing the step used in glycolysis because the reaction catalyzed by pyruvate kinase is irreversible under intracellular conditions. Pyruvate indeed has to take a rather tortuous route before being converted into the high energy phosphate compound phospho*enol*pyruvate. Ignoring momentarily some of the details, the conversion is accomplished by the coupling of two reactions requiring high energy phosphate compounds (an ATP and a GTP). The first is catalyzed by pyruvate carboxylase and the second by phospho*enol*pyruvate carboxykinase.





Since the GTP required for the phospho*enol*pyruvate carboxykinase catalyzed reaction is equivalent to an ATP through the action of nucleoside diphosphokinase (GDP + ATP $\rightleftharpoons$ GTP + ADP), and since CO_2 and HCO_3^- readily equilibrate by the action of carbonic anhydrase ($CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$), the sum of these reactions is

$$\text{Pyruvate}^- + 2\text{ATP}^{4-} \longrightarrow \text{phospho}\textit{enol}\text{pyruvate}^{3-} + 2\text{ADP}^{3-} + 2\text{P}_i^{2-} + 4\text{H}^+$$

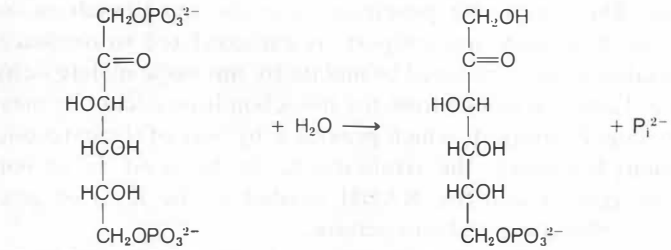
Whereas the conversion of phospho*enol*pyruvate to pyruvate by the enzyme pyruvate kinase is a highly favorable reaction, which yields the cell one molecule of ATP, the conversion of pyruvate into phospho*enol*pyruvate by the combination of pyruvate carboxylase and phos*enol*pyruvate carboxykinase is an equally unfavorable reaction, which costs the cell in effect 2 ATP molecules.

Now for some interesting details. As shown in Figure 7.14, the conversion of cytosolic pyruvate into cytosolic phospho*enol*pyruvate requires the participation of the mitochondrion. Pyruvate carboxylase is housed within the mitochondrion, making these particles mandatory for glucose synthesis. There are two routes that oxalacetate can then take to glucose—and both are important in human liver. This happens because phospho*enol*pyruvate carboxykinase occurs in both the cytosolic and mitosolic compartments. The simplest pathway to follow is the one involving the mitochondrial phospho*enol*pyruvate carboxykinase. In this case, oxalacetate is simply converted within the mitochondrion into phospho*enol*pyruvate, which then traverses the mitochondrial inner membrane in search of the rest of the enzymes of the gluconeogenic pathway. The second pathway would also be simple if oxalacetate could traverse the mitochondrial inner membrane to reach the cytosolic phospho*enol*pyruvate carboxykinase; however, as already discussed with respect to the malate–aspartate shuttle (Figure 7.5), oxalacetate per se cannot escape from the mitochondrion. Thus, the trick is again used, as in the malate–aspartate shuttle (Figure 7.5*B*) of converting oxalacetate into aspartate, which traverses the mitochondrial inner membrane by way of the aspartate–glutamate antiport. Aspartate is converted back to oxalacetate in the cytosol by transamination with α-ketoglutarate.

The steps from phospho*enol*pyruvate to fructose 1,6-bisphosphate are already familiar, being just the reverse of steps of the glycolytic pathway. The thing to note in this segment of the pathway is that the NADH generated by lactate dehydrogenase is utilized by the reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase.

6-Phosphofructokinase catalyzes an irreversible step in the

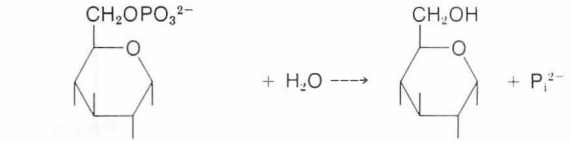
glycolytic pathway and is, therefore, of no use in the gluconeogenic pathway for the conversion of fructose 1,6-bisphosphate into fructose 6-phosphate. A way around this problem is offered by the enzyme fructose bisphosphatase, which catalyzes the irreversible reaction



Fructose 1,6-bisphosphate Fructose 6-phosphate

Note that ATP and ADP are not involved and that this is a reaction that can be used to yield fructose 6-phosphate, but, since it is irreversible, it is not a reaction that can be used in glycolysis to yield fructose 1,6-bisphosphate.

The reaction catalyzed by phosphoglucoisomerase is freely reversible and functions in both the glycolytic and gluconeogenic pathways. However, glucose 6-phosphatase has to be used instead of glucokinase for the last step. Glucose 6-phosphatase catalyzes an irreversible reaction under intracellular conditions.



α-D-Glucose 6-phosphate α-D-Glucose

It should be noted again that nucleotides do not have a role in this reaction and that the function of this enzyme is to generate glucose, not to convert glucose into glucose 6-phosphate.

Special Features of Gluconeogenesis

The pathway for lactate gluconeogenesis (Figure 7.14) illustrates that, regardless of whether carbon exits the mitochondrion in the form of phospho*enol*pyruvate or aspartate, there is a perfect coupling between the NADH generated by lactate dehydrogenase and the NADH used by glyceraldehyde 3-phosphate dehydrogenase. This is tidy—none is left over and no extra is required. Consider pyruvate, however, as a gluconeogenic substrate (Figure 7.15). Pyruvate

should, one would think, just follow the same pathway as the pyruvate generated from lactate by lactate dehydrogenase as in Figure 7.14. The NADH needed by the glyceraldehyde 3-phosphate dehydrogenase-catalyzed reaction, however, would not be generated by such a pathway. The problem is solved nicely, as shown in Figure 7.15, by having the carbon exit from the mitochondrion as malate. Thus, pyruvate penetrates into the mitochondrion on the pyruvate–hydroxide ion antiport, is carboxylated to oxaloacetate, and oxaloacetate is reduced to malate by mitosolic malate dehydrogenase. Egress of malate from the mitochondrion occurs by means of the malate–P_i antiport, which provides, by way of the cytosolic malate dehydrogenase, the oxaloacetate to be used as carbon for gluconeogenesis and the NADH needed at the level of glyceraldehyde 3-phosphate dehydrogenase.

All amino acids except leucine and lysine can supply carbon which can be used for the net synthesis of glucose by gluconeogenesis. The details of the pathways of amino acid catabolism are covered in Chapter 12. For our purposes here it is important to note only that if the catabolism of an amino acid can yield either net pyruvate or net oxalacetate formation, then net glucose synthesis can occur from that amino acid. As in Figures 7.14 and 7.15, oxalacetate is an intermediate in gluconeogenesis and pyruvate is readily converted to oxalacetate by the action of pyruvate carboxylase. The abbreviated pathway, given in Figure 7.12,

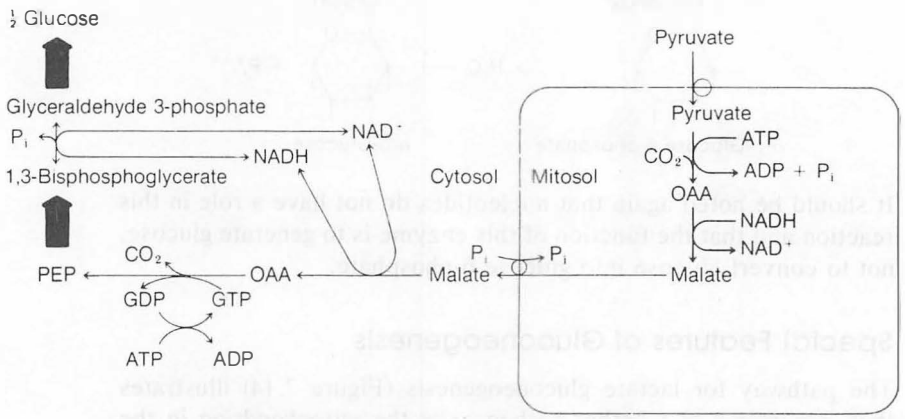


Figure 7.15
The pathway of gluconeogenesis from pyruvate.
Abbreviations are as in Figure 7.14. Large arrows indicate portions of the pathway which are identical to those given in Figure 7.14.

illustrates how amino acid catabolism fits with the process of gluconeogenesis. The catabolism of amino acids feeds carbon into the TCA cycle at more than one point (see Figure 12.28 of Chapter 12). However, as long as net synthesis of a TCA cycle intermediate occurs as a consequence of the catabolism of a particular amino acid, net synthesis of oxalacetate will have to follow. Reactions which "fill up" the TCA cycle with intermediates, that is, lead to the net synthesis of TCA cycle intermediates, are called anaplerotic reactions. Such reactions support gluconeogenesis because they provide for the net synthesis of oxalacetate. The reactions catalyzed by pyruvate carboxylase and glutamate dehydrogenase are good examples of anaplerotic reactions (anaplerosis):

$$\text{Pyruvate}^- + \text{ATP}^{4-} + \text{HCO}_3^- \longrightarrow \text{oxalacetate}^{2-} + \text{ADP}^{3-} + \text{P}_i^{2-} + \text{H}^+$$

$$\text{Glutamate}^- + \text{NAD(P)}^+ \longrightarrow \alpha\text{-ketoglutarate}^{2-} + \text{NAD(P)H} + \text{NH}_4^+ + \text{H}^+$$

On the other hand, the reaction catalyzed by glutamate-oxalacetate transaminase is not an anaplerotic reaction,

$$\alpha\text{-Ketoglutarate} + \text{aspartate} \rightleftharpoons \text{glutamate} + \text{oxalacetate}$$

because net synthesis of a TCA cycle intermediate is not accomplished (note presence of an intermediate of the TCA cycle on both sides of the equation).

Glucose Synthesis from Amino Acids

Gluconeogenesis from amino acids imposes an additional nitrogen load upon the liver, as pointed out in the description of the alanine cycle in Figure 7.13*b*. Since the liver has to convert this nitrogen into urea, there is a close relationship between urea synthesis and glucose synthesis from amino acids. This relationship is illustrated in Figure 7.16 for alanine, the most important gluconeogenic amino acid. Two alanine molecules are shown being transaminated to give two molecules of pyruvate, which enter the mitochondrion where each is used by a separate pathway, one pathway yielding malate plus NH_4^+ and the other aspartate. The latter two nitrogen-containing compounds should be recognized as the primary substrates for the urea cycle. The aspartate leaves the mitochondrion and becomes part of the urea cycle after reacting with citrulline. The carbon of aspartate is released from the urea cycle in the form of fumarate, the latter then being converted to malate by

cytosolic fumarase. Both this malate and the malate exiting from the mitochondria are converted to glucose by the action of the cytosolic enzymes of the gluconeogenic pathway. As shown in Figure 7.16, a balance is achieved between the reducing equivalents (NADH) generated and those required in both the cytosolic and mitosolic spaces.

Leucine and lysine are the only amino acids that cannot function as carbon sources for the net synthesis of glucose. These are the only amino acids that are only ketogenic and not also glucogenic. As shown in Table 7.2, all other amino acids are classified as glucogenic, or at least both glucogenic and ketogenic. Glucogenic amino acids give rise to the net synthesis of either pyruvate or oxalacetate, whereas glucogenic-ketogenic amino acids also yield the ketone body acetoacetate, or at least acetyl CoA which is readily converted into ketone bodies. Acetyl CoA is the end product of lysine metabolism, and acetoacetate and acetyl CoA are the end products of leucine metabolism. In the human and other animals, no pathway exists for converting acetoacetate or acetyl CoA into pyru-

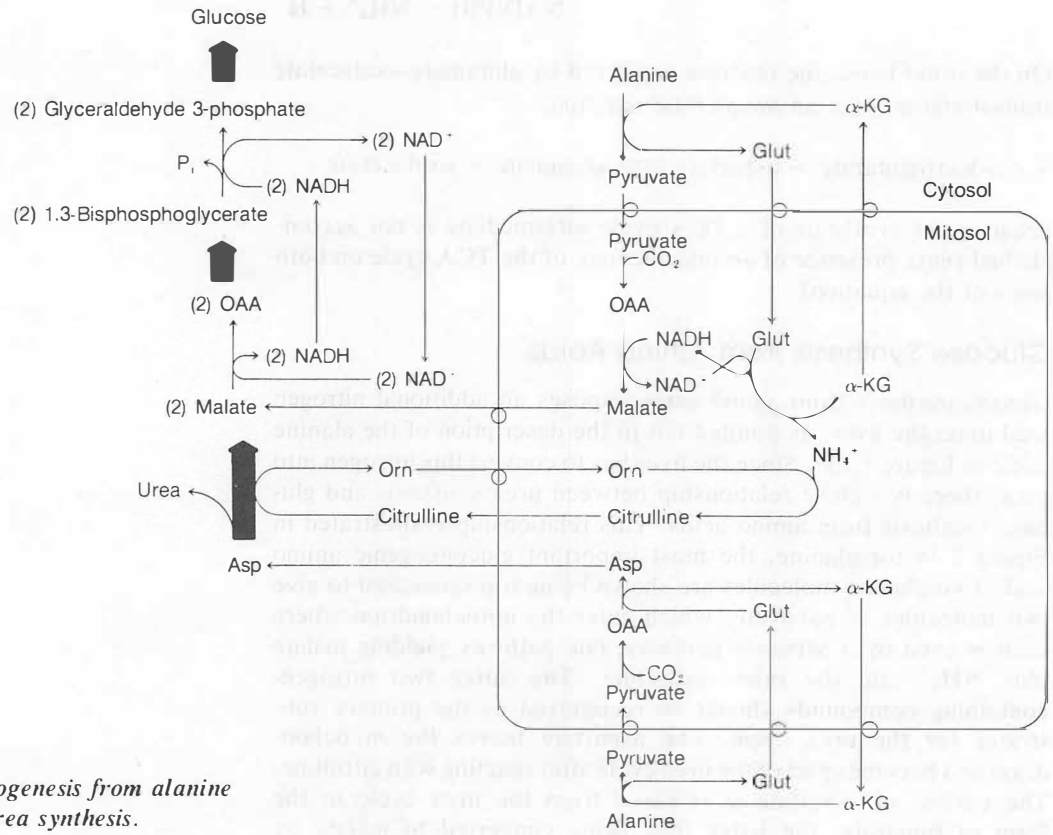


Figure 7.16
The pathway of gluconeogenesis from alanine and its relationship to urea synthesis.

vate or oxalacetate. It may not be immediately obvious why acetyl CoA cannot be used for net synthesis of glucose but remember that the reaction catalyzed by pyruvate dehydrogenase is irreversible:

$$\text{Pyruvate}^- + \text{NAD}^+ + \text{CoASH} \longrightarrow \text{acetyl CoA} + \text{NADH} + \text{CO}_2$$

meaning this reaction cannot be used to synthesize pyruvate from acetyl CoA. It might be argued the oxaloacetate is generated from acetyl CoA by way of the TCA cycle:

$$\text{Acetyl CoA} \longrightarrow \text{citrate} \xrightarrow{\text{TCA}} 2\text{CO}_2 + \text{oxalacetate}$$

However, this is a fallacious argument because oxalacetate must react with acetyl CoA to give citrate by way of citrate synthase:

$$\text{Acetyl CoA} + \text{oxalacetate} \longrightarrow \text{citrate} + \text{CoA}$$

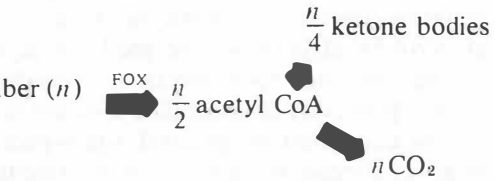
$$\text{Citrate} \xrightarrow{\text{TCA}} 2\text{CO}_2 + \text{oxalacetate}$$

$$\text{Sum:} \quad \text{Acetyl CoA} \longrightarrow 2\text{CO}_2 + \text{CoA}$$

The point is that, although students of biochemistry have tried every conceivable way in the laboratory and on examinations, it turns out to be impossible for animals to synthesize net oxalacetate or glucose from acetyl CoA.

Glucose Synthesis from Fat

This lack of an anaplerotic pathway from acetyl CoA also means that in general it is impossible to synthesize glucose from fatty acids. Most fatty acids found in the human body are of the straight-chain variety with an even number of carbon atoms. Their catabolism by fatty acid oxidation (FOX) followed by ketogenesis or complete oxidation to CO_2 can be abbreviated:



Since acetyl CoA and other intermediates of fatty acid oxidation cannot be converted to oxalacetate or any other intermediate of gluconeogenesis, it is impossible to synthesize glucose from fatty

Table 7.2 The Glucogenic and Ketogenic Amino Acids

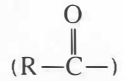
Glucogenic	*Ketogenic*	*Both*
Glycine	Leucine	Threonine
Serine	Lysine	Isoleucine
Valine		Phenylalanine
Histidine		Tyrosine
Arginine		Tryptophan
Cysteine		
Proline		
Hydroxyproline		
Alanine		
Glutamate		
Glutamine		
Aspartate		
Asparagine		
Methionine		

acids. An exception to this general rule applies to fatty acids with methyl branches (e.g., phytanic acid, obtained as a breakdown product of chlorophyll; see discussion of Refsum's disease, Clin. Corr. 9.4) and fatty acids with an odd number of carbon atoms. The catabolism of such compounds yields propionyl CoA:

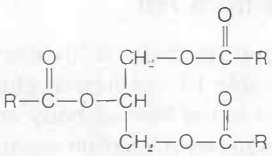
Fatty acid with odd number (n) of carbon atoms $\longrightarrow$ $\frac{(n-3)}{2}$ acetyl CoA + 1 propionyl CoA

Propionate is a good precursor for gluconeogenesis, generating oxalacetate by the anaplerotic pathway shown in Figure 7.17. Although not indicated in the figure, all of the steps involved in anaplerosis from propionate take place within the mitochondrion, and carbon has to exit from the mitochondrion in the form of malate in order to balance the reducing equivalent stoichiometry during gluconeogenesis. Propionate is also produced in the catabolism of valine and isoleucine and the conversion of cholesterol into bile acids.

It is sometimes loosely stated that fat *cannot* be converted into carbohydrate (glucose) by the liver. In a sense this is certainly true, that is, fatty acid metabolism, with the exception of fatty acids with branched chains or an odd number of carbon atoms, cannot give rise to net synthesis of glucose. However, the term "fat" is usually used to refer to triacylglycerols—and a triacylglycerol is composed of three *O*-acyl groups combined with 1 glycerol molecule:



Acyl group



Hydrolysis of this molecule of fat yields three fatty acids and glycerol, the latter compound being an excellent substrate for gluconeogenesis as shown in Figure 7.18. Phosphorylation of glycerol by glycerol kinase produces glycerol 3-phosphate, which can be converted back into fat by esterification with fatty acyl CoA esters. However, of immediate concern is that glycerol 3-phosphate can be converted by glycerol 3-phosphate dehydrogenase into dihydroxyacetone phosphate, an intermediate of the gluconeogenic pathway (see Figure 7.14). As also indicated in Figure 7.18, the last stage of glycolysis can compete with the gluconeogenic pathway and convert dihydroxyacetone phosphate into lactate (or into pyruvate for subsequent complete oxidation to CO_2 and H_2O).

Glucose Synthesis from Other Sugars

Fructose

Fructose is also an excellent substrate for gluconeogenesis. Humans consume considerable quantities of fructose in the form of sucrose, and much of the fructose obtained by sucrose hydrolysis in the small bowel is converted into glucose in the liver by gluconeogenesis. Like glucose, fructose is phosphorylated in the liver by a special ATP-linked kinase (Figure 7.19). Phosphorylation of fructose occurs in

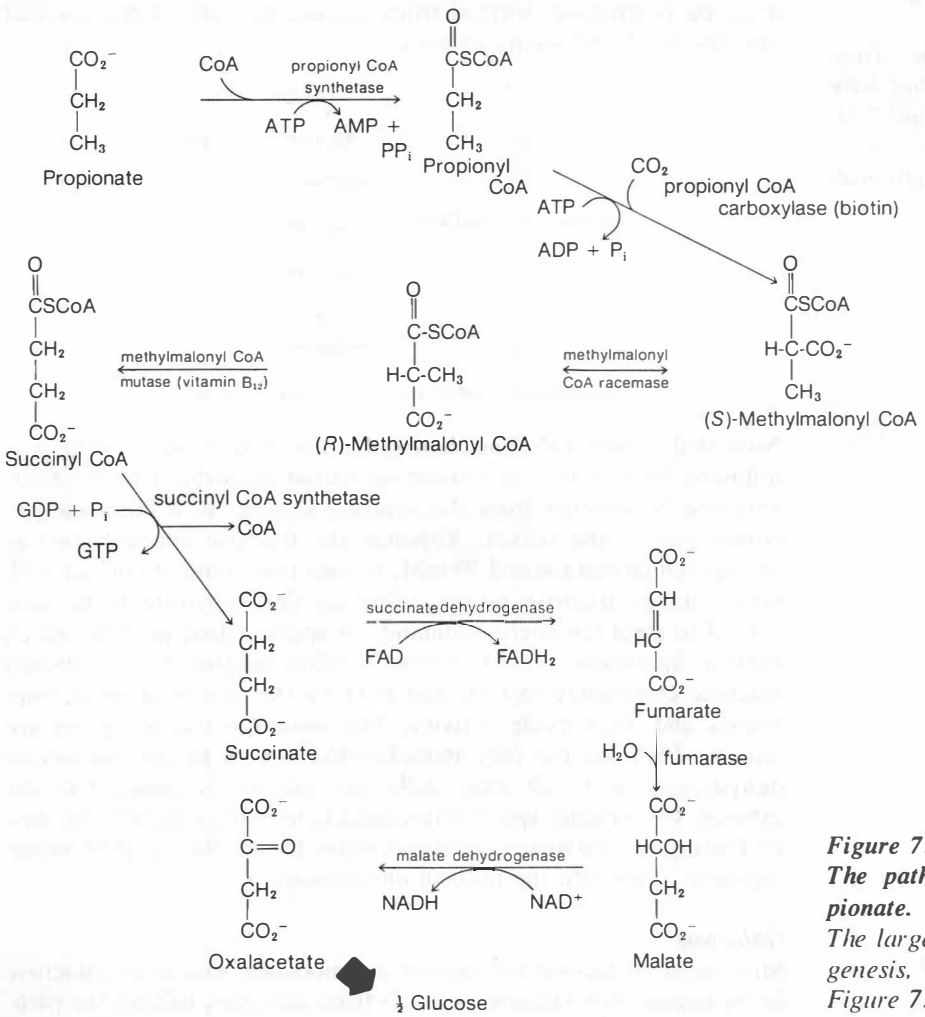


Figure 7.17
The pathway of gluconeogenesis from propionate.
The large arrow refers to steps of gluconeogenesis, which have been given in detail in Figure 7.14.

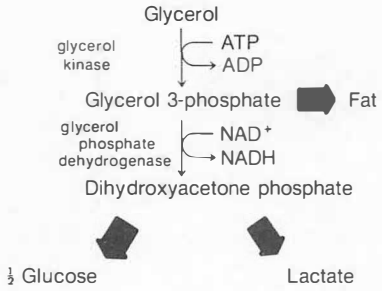
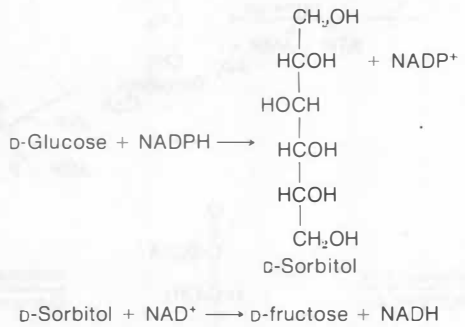


Figure 7.18
The pathway of gluconeogenesis from glycerol, along with competing pathways.
Large arrows indicate steps of the glycolytic and gluconeogenic pathways that have been given in detail in Figures 7.3 and 7.14, respectively. The large arrow pointing to fat refers to the synthesis of triacylglycerols and glycerophospholipids.

the 1 position to yield fructose 1-phosphate (see Clin. Corr. 7.3). A special aldolase then cleaves fructose 1-phosphate to give one molecule of dihydroxyacetone phosphate and one molecule of glyceraldehyde. The latter compound can be reduced to glycerol and used by the same pathway given for glycerol in the prior figure. The two molecules of dihydroxyacetone phosphate eventually obtained from one molecule of fructose can then be converted to glucose by enzymes of the gluconeogenic pathway or, alternatively, into pyruvate or lactate by the last stage of glycolysis. In analogy to glycolysis, the conversion of fructose into lactate is termed fructolysis.

Besides its dietary role, fructose is also generated in the body of man for an interesting purpose. The major energy source of spermatozoa is fructose, formed from glucose by cells of the seminal vesicles by the following pathway:



Note that a NADPH-dependent reduction of glucose to sorbitol is followed by a NAD^+-dependent oxidation of sorbitol to fructose. Fructose is secreted from the seminal vesicles in a fluid that becomes part of the semen. Although the fructose concentration in human semen can exceed 10 mM, tissues that come in contact with semen utilize fructose poorly, allowing this substrate to be conserved to meet the energy demands of spermatozoa in their search for ova. Spermatozoa contain mitochondria and thus can metabolize fructose completely to CO_2 and H_2O by the combination of fructolysis and TCA cycle activity. The mitochondria of sperm are unique. They are the only mitochondria known to contain lactate dehydrogenase. In all other cells this enzyme is confined to the cytosol. This enables sperm mitochondria to oxidize lactate obtained by fructolysis and makes shuttle systems for the transport of reducing equivalents into the mitosol unnecessary.

Galactose

Milk sugar or lactose constitutes an important source of galactose in the human diet. Gluconeogenesis from galactose follows the pathway shown in Figure 7.20. The role of UDP-glucose as a recycling

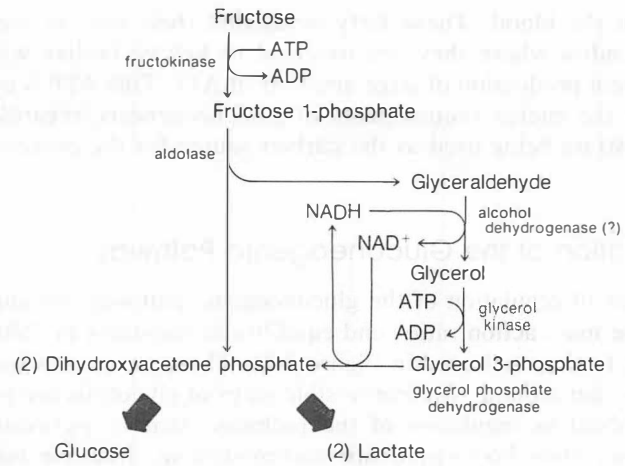


Figure 7.19
The pathway of gluconeogenesis from fructose, along with the competing pathway of fructolysis.
Large arrows indicate steps of the glycolytic and gluconeogenic pathways that have been given in detail in Figures 7.3 and 7.14, respectively.

intermediate in the overall process of converting galactose into glucose should be noted. The absence of the enzyme galactose 1-phosphate uridylyltransferase accounts for most cases of galactosemia (see Clin. Corr. 8-3).

Mannose

Mannose is found in our diet, but in very limited quantities. Fortunately, its pathway of metabolism is also short and simple. It is first phosphorylated by hexokinase and then converted into fructose 6-phosphate by mannose phosphate isomerase:

$$\text{D-Mannose} + \text{ATP} \longrightarrow \text{D-mannose 6-phosphate} + \text{ADP}$$

$$\text{D-Mannose 6-phosphate} \rightleftharpoons \text{D-fructose 6-phosphate}$$

As we already know, the latter compound can then be used in either the glycolytic pathway or the gluconeogenic pathway.

Cost of Glucose Synthesis

The synthesis of glucose is costly in terms of ATP, for example, at least 6 molecules of ATP are required for the synthesis of 1 molecule of glucose from 2 molecules of lactate. The ATP needed by the liver cell for glucose synthesis is provided in large part by fatty acid oxidation. Metabolic conditions under which the liver is required to synthesize glucose generally favor increased availability of fatty

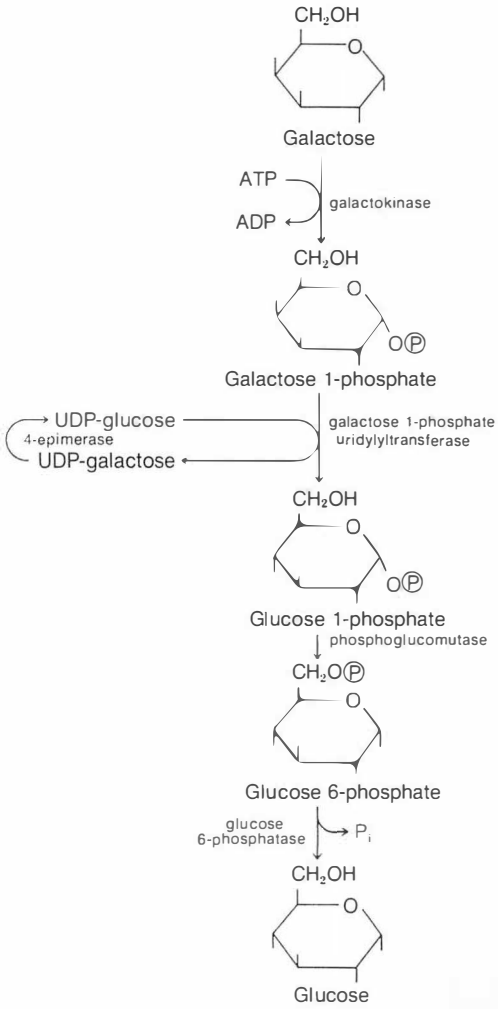


Figure 7.20
The pathway of gluconeogenesis from galactose.

acids in the blood. These fatty acids find their way to the liver mitochondria where they are oxidized to ketone bodies with the concurrent production of large amounts of ATP. This ATP is used to support the energy requirements of gluconeogenesis, regardless of the substrate being used as the carbon source for the process.

Regulation of the Gluconeogenic Pathway

The sites of regulation of the gluconeogenic pathway are apparent from the mass action ratios and equilibrium constants in Table 7.1, and are further indicated in Figure 7.21. Those enzymes which are used to "get around" the irreversible steps of glycolysis are primarily involved in regulation of the pathway, that is, pyruvate carboxylase, phospho*enol*pyruvate carboxykinase, fructose bisphosphatase, and glucose 6-phosphatase. Considering the regulation of hepatic gluconeogenesis is almost the same as considering the regulation of hepatic glycolysis, which was discussed in some detail in earlier sections of this chapter. Inhibition of glycolysis at its chief regulatory sites, or repressing the synthesis of the enzymes involved at these sites (glucokinase, phosphofructokinase, and pyruvate kinase), greatly increases the effectiveness of the opposing gluconeogenic enzymes. Turning on gluconeogenesis is accomplished in large part, therefore, by shutting off glycolysis. Fatty acid oxidation does more than just supply ATP for the process because it actually promotes glucose synthesis. First of all, it increases the steady-state concentration of mitochondrial acetyl CoA, a positive allosteric effector of the reaction catalyzed by the mitochondrial enzyme pyruvate carboxylase. Second, the increase in acetyl CoA and pyruvate carboxylase activity results in a greater synthesis of citrate, a negative effector of phosphofructokinase. A secondary effect of inhibition of phosphofructokinase is a decrease in fructose 1,6-bisphosphate concentration, an activator of pyruvate kinase. This decreases the flux of phospho*enol*pyruvate to pyruvate by pyruvate kinase, and increases the effectiveness of the combined efforts of pyruvate carboxylase and phospho*enol*pyruvate carboxykinase in the conversion of pyruvate to phospho*enol*pyruvate. An increase in ATP levels with the consequential decrease in AMP levels would favor gluconeogenesis by way of inhibition of phosphofructokinase and pyruvate kinase and activation of fructose bisphosphatase (see Figure 7.21 and the discussion of the regulation of glycolysis, page 348). A shortage of oxygen for respiration, a shortage of fatty acids for oxidation, or any inhibition or uncoupling of oxidative phosphorylation would be expected to cause the liver to turn from gluconeogenesis to glycolysis.

Hormonal control of gluconeogenesis is a matter of regulating the supply of fatty acids to the liver and, in addition, regulating the enzymes of both the glycolytic and gluconeogenic pathways. Glucagon increases plasma fatty acids by promoting lipolysis in adipose tissue, an action which is opposed by insulin. The greater availability of fatty acids caused by glucagon results in more fatty acid oxidation by the liver which, as discussed above, promotes glucose synthesis. Insulin, on the other hand, has the opposite effect. Glucagon and insulin also regulate gluconeogenesis in the liver by influencing the state of phosphorylation of the interconvertible enzymes of the liver. As discussed above (page 358), two enzymes of the glycolytic pathway are subject to interconversion, that is, subject to regulation by the phosphorylation–dephosphorylation mechanism. Pyruvate kinase and 6-phosphofructokinase both appear to be active in the dephosphorylated mode and inactive in the phosphorylated mode. cAMP is the signal that appears to promote phosphorylation of these enzymes (see Chapters 4 and 14, and the regulation of glycogen metabolism, page 390, for a more detailed discussion of interconvertible enzymes). Glucagon activates adenylate cyclase to produce cyclic AMP, which activates protein kinase, which, in turn, causes the phosphorylation and inactivation of pyruvate kinase and 6-phosphofructokinase. Inhibition of these enzymes greatly acceler-

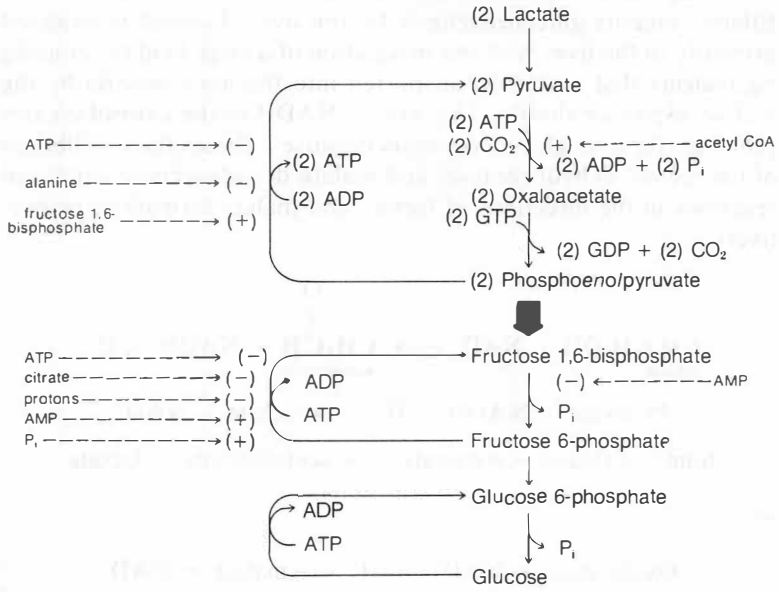


Figure 7.21
Important allosteric regulatory features of the gluconeogenic pathway.

CLIN. CORR. 7.10 HYPOGLYCEMIA AND ALCOHOL INTOXICATION

Consumption of alcohol, especially by an undernourished person, can cause hypoglycemia. The same effect can result from drinking alcohol after strenuous exercise. In both cases the hypoglycemia results from the inhibitory effects of alcohol on hepatic gluconeogenesis and thus occurs under circumstances of hepatic glycogen depletion. The problem is caused by the NADH produced during the metabolism of alcohol. The liver simply cannot handle the reducing equivalents provided by ethanol oxidation fast enough to prevent metabolic derangements. The extra reducing equivalents block the conversion of lactate to glucose and promote the conversion of alanine into lactate, resulting in considerable lactate accumulation in the blood. Since lactate has no place to go, lactic acidosis (see Clin. Corr. 7.5) can develop.

Low doses of alcohol cause impaired motor and intellectual performance; high doses have a depressant effect, which can lead to stupor and anesthesia. Low blood sugar can contribute to these undesirable effects of alcohol. What is more, a patient may be thought to be simply inebriated when in fact he is suffering hypoglycemia, with irreversible damage to the central nervous system. Alcohol-induced hypoglycemia has also been suggested to be a causative factor in the fetal alcohol syndrome. This syndrome, characterized by growth retardation and abnormalities of the central nervous system, is seen in the offspring of alcoholic and heavy-drinking women, and may be produced by even "social" drinking during pregnancy.

ates gluconeogenesis by blocking futile cycling of substrates back through these portions of the glycolytic pathway. Insulin has effects opposite to those of glucagon—but the nature of the intracellular messenger formed in response to insulin is not known at this time.

Glucagon and insulin also have long-term effects upon the levels of hepatic enzymes involved in glycolysis and gluconeogenesis. A high glucagon : insulin ratio in the blood increases the capacity for gluconeogenesis and decreases the capacity for glycolysis in the liver. A low glucagon : insulin ratio has the opposite effects. In addition to the short-term or acute mechanisms discussed above, this is accomplished by induction and repression of the synthesis of key enzymes of the pathways. Thus the glucagon : insulin ratio in the blood increases when gluconeogenesis is needed. This serves to signal the induction within the liver of the synthesis of greater quantities of phospho*enol*pyruvate carboxykinase, fructose bisphosphatase, and glucose 6-phosphatase. The same signal causes the repression of the synthesis of glucokinase, 6-phosphofructokinase, and pyruvate kinase. The opposite response occurs when glucose synthesis is not needed, that is, when a low glucagon : insulin ratio prevails because of maintenance of high blood glucose levels by glucose input from the gastrointestinal tract.

It is impossible to leave gluconeogenesis without saying something about the effects of alcohol on the process (see Clin. Corr. 7.10). Ethanol inhibits gluconeogenesis by the liver. Ethanol is oxidized primarily in the liver with the production of a large load of reducing equivalents that must be transported into the mitochondria by the malate–aspartate shuttle. This excess NADH in the cytosol creates problems for liver gluconeogenesis because it forces the equilibrium of the lactate dehydrogenase- and malate dehydrogenase-catalyzed reactions in the directions of lactate and malate formation, respectively:

$$\underset{\text{Ethanol}}{CH_3CH_2OH} + NAD \longrightarrow \underset{\text{Acetaldehyde}}{CH_3\overset{\overset{O}{\|}}{C}H} + NADH + H^+$$

$$\text{Pyruvate} + NADH + H^+ \longrightarrow \text{lactate} + NAD^+$$

$$\text{Sum: Ethanol} + \text{pyruvate} \longrightarrow \text{acetaldehyde} + \text{lactate}$$

or

$$\text{Oxalacetate} + NADH + H^+ \longrightarrow \text{malate} + NAD^+$$

$$\text{Sum: Ethanol} + \text{oxalacetate} \longrightarrow \text{acetaldehyde} + \text{malate}$$

In the presence of ethanol there is no shortage of NADH for the

gluconeogenic pathway at the level of glyceraldehyde 3-phosphate dehydrogenase; however, forcing the equilibrium of lactate dehydrogenase and malate dehydrogenase as shown above inhibits glucose synthesis because pyruvate and oxalacetate are no longer available in sufficient concentrations for the reactions catalyzed by pyruvate carboxylase and phospho*enol*pyruvate carboxykinase, respectively. The take home message is: Don't drink while glucose synthesizing!

7.6 GLYCOGENOLYSIS AND GLYCOGENESIS

Significance

Glycogenolysis refers to the intracellular breakdown of glycogen; glycogenesis to the intracellular synthesis of glycogen. We will be concerned here mainly with these processes in muscle and liver because of their greater quantitative importance in these tissues. However, it should be appreciated that these processes are of some importance in almost every tissue of the body.

The liver has tremendous capacity for storing glycogen. In the well-fed human the liver glycogen content can account for as much as 10% of the wet weight of this organ. Muscle stores less when expressed on the same basis—a maximum of only 1–2% of its wet weight. However, since the average person has more muscle than liver, there is about twice as much total muscle glycogen as liver glycogen.

Muscle and liver glycogen stores serve completely different roles. Muscle glycogen is present to serve as a fuel reserve for the synthesis of ATP within that tissue, whereas liver glycogen functions as a glucose reserve for the maintenance of blood glucose concentrations. Liver glycogen levels vary greatly in response to the intake of food, accumulating to high levels shortly after a meal and then decreasing slowly as it is mobilized to help maintain a nearly constant blood glucose level (see Figure 7.22). Liver glycogen reserves in the human are called into play between meals and to an even greater extent during the nocturnal fast. In both man and the rat, the store of glycogen in the liver lasts somewhere between 12 and 24 h during fasting, depending greatly, of course, upon whether the individual under consideration is caged or running wild.

Glycogen in muscle is used within this tissue when needed as a source of ATP for increased muscular activity. Most of the glucose of the glycogen molecule is consumed within muscle cells without

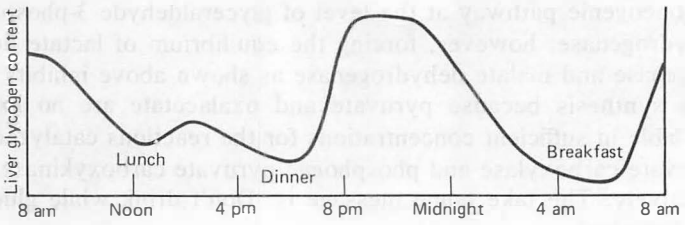


Figure 7.22
Variation of liver glycogen levels between meals and during the nocturnal fast.

the formation of free glucose as an intermediate. However, because of a peculiarity of glycogen catabolism to be discussed below, about 8% of muscle glycogen is converted into free glucose within the tissue. Some of this glucose is released into the bloodstream, but most gets metabolized by the glycolytic pathway (Figure 7.3) in the muscle. Since muscle cells lack glucose 6-phosphatase, and since most of the free glucose formed during glycogen breakdown is further catabolized, muscle glycogen is not of quantitative importance in the maintenance of blood glucose levels. Muscle glycogen levels vary much less than liver glycogen levels in response to food intake. The processes of glycogenesis and glycogenolysis within the liver work to "buffer" blood glucose levels, but this is not an important role of these processes in muscle. Exercise of a muscle, that is, increased mechanical work, is what triggers the mobilization of muscle glycogen. It serves to supply the ATP needed for the performance of work. The yield of ATP and the fate of the carbon of glycogen depend upon whether a "white" or "red" muscle is under consideration. Red muscle fibers are supplied with a rich blood flow, contain large amounts of myoglobin, and are packed with mitochondria. Glycogen mobilized within these cells is converted into pyruvate, which, because of the availability of O_2 and mitocondria, can be converted into CO_2 and H_2O. In contrast, white muscle fibers have a poorer blood supply and fewer mitochondria. Glycogenolysis within this tissue supplies substrate for glycolysis, with the end

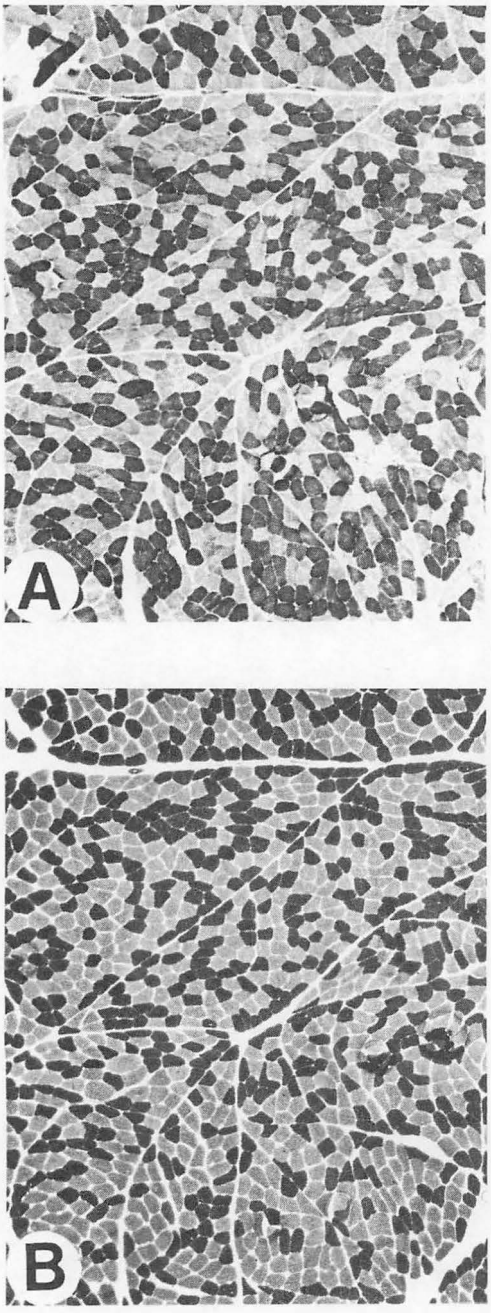
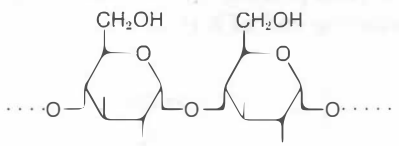


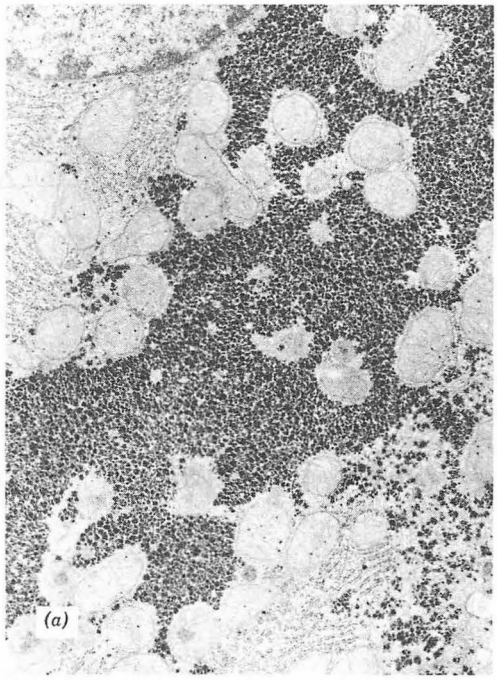
Figure 7.23
Cross section of human skeletal muscle showing red and white muscle fibers.
Sections were stained for NADH diaphorase activity in A, for ATPase activity at pH 9.4 in B. The red fibers are dark and the white fibers are light in A; vice versa in B.
Pictures provided by Dr. Michael H. Brooke of the Jerry Lewis Neuromuscular Research Center, St. Louis, Mo.

product being primarily lactate. White muscle fibers have enormous capacity for glycogenolysis and glycolysis, much more than red muscle fibers. Since their glycogen stores are limited, however, muscles of this type can only function at full capacity for relatively short periods of time. Breast muscle and the heart of chicken are good examples of white and red muscles, respectively. The heart has to beat continuously and is therefore blessed with many mitochondria and a rich supply of blood via the coronary arteries. The heart stores glycogen to be used when a greater work load is imposed. The breast muscle of the chicken, in contrast to the heart, is not continuously carrying out work. Its important function is to enable the chicken to fly rapidly for short distances, as in fleeing from predators (or bully roosters). Because glycogen can be mobilized so rapidly, these muscles are designed for maximal activity for a relatively short period of time. Although it was easier to point out readily recognizable white and red muscles in the chicken, most skeletal muscles of the human body are composed of a mixture of red and white fibers in order to provide for both rapid and sustained muscle activity. The distribution of white and red muscle fibers in cross sections of a human skeletal muscle can be readily shown by using special staining procedures (see Figure 7.23).

Glycogen granules are abundant in the liver of the well-fed animal but are virtually absent from the liver of the 24-hour-fasted animal (Figure 7.24). Heavy exercise causes the same loss of glycogen granules in muscle fibers. These granules of glycogen correspond to clusters of glycogen molecules, the molecular weights of which can approach 2×10^7. The structure of glycogen is interesting in that it is a polymeric structure with innumerable branches. Glycogen is composed entirely of glucosyl residues, the majority of which are linked together by α-[1 → 4] glycosidic linkages:



α-[1 → 4] linkage



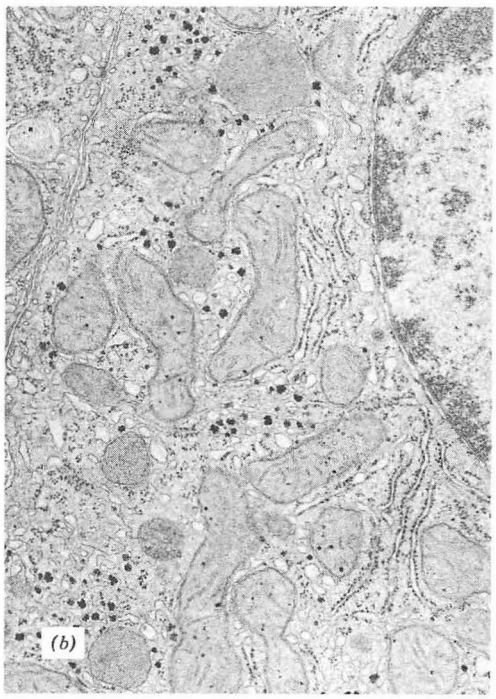


Figure 7.24

Electron micrographs showing glycogen granules (darkly stained material) in the liver of a well-fed rat (a) and the relative absence of such granules in the liver of a rat starved for 24 h (b).

Micrographs provided by Dr. Robert R. Cardell of the Department of Anatomy at the University of Cincinnati.

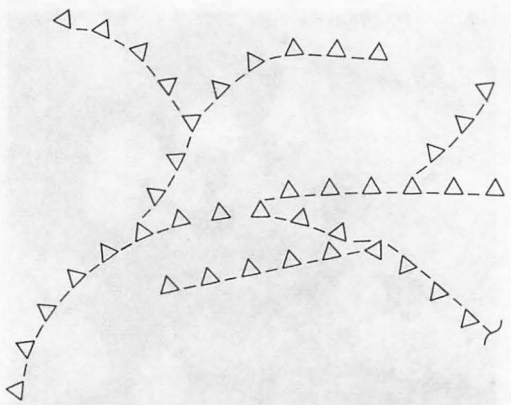
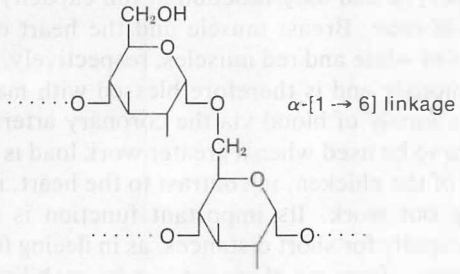


Figure 7.25
The branched structure of glycogen.

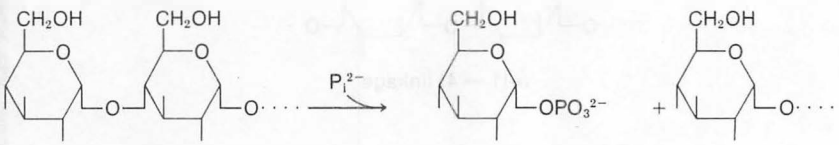
Branches occur in the glycogen molecule because of frequent α-[1 → 6] glycosidic linkages:



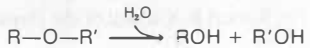
A limb of the glycogen "tree" (see Figure 7.25) is characterized by branches at every fourth glucosyl residue witihin the more central core of the molecule. These branches occur much less frequently in the outer tiers of the molecule. An interesting question, which we shall attempt to answer below, is why this polymer is constructed by the cell with so many intricate branches and loose ends? Glycogen certainly stands in contrast to proteins and nucleic acids in this regard but, of course, it is a storage form of fuel and never has to catalyze a reaction nor convey information within a cell.

Pathway of Glycogen Degradation

The first step of glycogen degradation is catalyzed by the enzyme glycogen phosphorylase (see Figure 7.26). This enzyme catalyzes the phosphorolysis of glycogen, a reaction in which the elements of inorganic phosphate are used in the cleavage of an α-[1 → 4] glycosidic bond to yield glucose 1-phosphate. This always occurs at a terminal, nonreducing end of a glycogen molecule:



Phosphorolysis should be contrasted with hydrolysis, in which the elements of water are used in a cleavage reaction:



Inorganic phosphate is used in a phosphorolytic cleavage reaction:

$$R{-}O{-}R' \xrightarrow{P_i^{2-}} R{-}O{-}PO_3^{2-} + R'OH$$

Since a molecule of glycogen may contain up to 100,000 glucose residues, its structure is usually abbreviated $(glucose)_n$. The reaction catalyzed by the enzyme glycogen phosphorylase can then be written as

$$(Glucose)_n + P_i^{2-} \longrightarrow (glucose)_{n-1} + \alpha\text{-D-glucose 1-phosphate}^{2-}$$

The next step of glycogen degradation is catalyzed by phosphoglucomutase:

$$\text{Glucose 1-phosphate} \rightleftharpoons \text{glucose 6-phosphate}$$

This is a near-equilibrium reaction under intracellular conditions, allowing it to function in both glycogen degradation and synthesis. It has the interesting feature of having a reaction mechanism analogous to that catalyzed by phosphoglyceromutase (page 340) in that a bisphosphate compound is an obligatory intermediate:

$$\text{E-P + glucose 1-phosphate} \rightleftharpoons \text{E + glucose 1,6-bisphosphate}$$

$$\text{E + glucose 1,6-bisphosphate} \rightleftharpoons \text{E-P + glucose 6-phosphate}$$

$$\text{Sum: glucose 1-phosphate} \rightleftharpoons \text{glucose 6-phosphate}$$

As with phosphoglyceromutase, a catalytic amount of the bisphosphate compound must be present for the reaction to occur. It is produced in small quantities for this specific purpose by an enzyme called phosphoglucokinase:

$$\text{Glucose 6-phosphate + ATP} \longrightarrow \text{glucose 1,6-bisphosphate + ADP}$$

The next enzyme involved in glycogenolysis depends on the tissue under consideration (see Figure 7.26). In liver the glucose 6-phosphate produced by glycogenolysis would be primarily hydrolyzed by glucose 6-phosphatase to give free glucose:

$$\text{Glucose 6-phosphate}^{2-} + H_2O \longrightarrow \text{glucose} + P_i^{2-}$$

Lack of this enzyme results in type I glycogen storage disease (see Clin. Corr. 7.11). The overall, balanced equation for the removal of

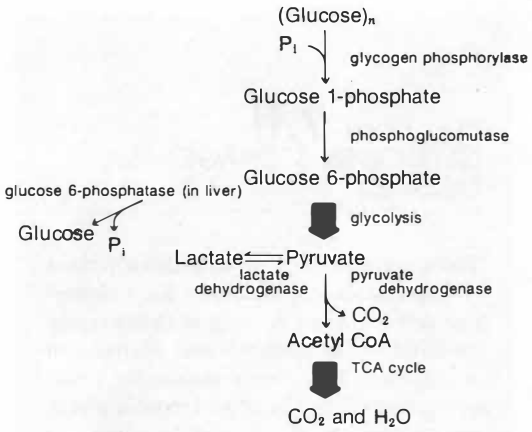


Figure 7.26
Glycogenolysis and the fate of glycogen degraded in liver versus its fate in peripheral tissues.
Large arrows refer to glycolysis and TCA activity.

one glucosyl residue from glycogen in the liver by glycogenolysis is then

$$(\text{Glucose})_n + H_2O \longrightarrow (\text{glucose})_{n-1} + \text{glucose}$$

In other words, glycogenolysis in the liver involves phosphorolysis but, because the phosphate ester is cleaved by a phosphatase, the overall reaction adds up to be hydrolysis of glycogen. It should be noted that no ATP is used or formed in the process of glycogenolysis.

In peripheral tissues the glucose 6-phosphate generated by glycolysis would be used by the glycolytic pathway, which would lead primarily to the generation of lactate in white muscle and primarily to the complete oxidation of the glucose carbon to CO_2 and H_2O in red muscle. Since no ATP had to be invested to phosphorylate the glucose 6-phosphate obtained from glycogen, the overall

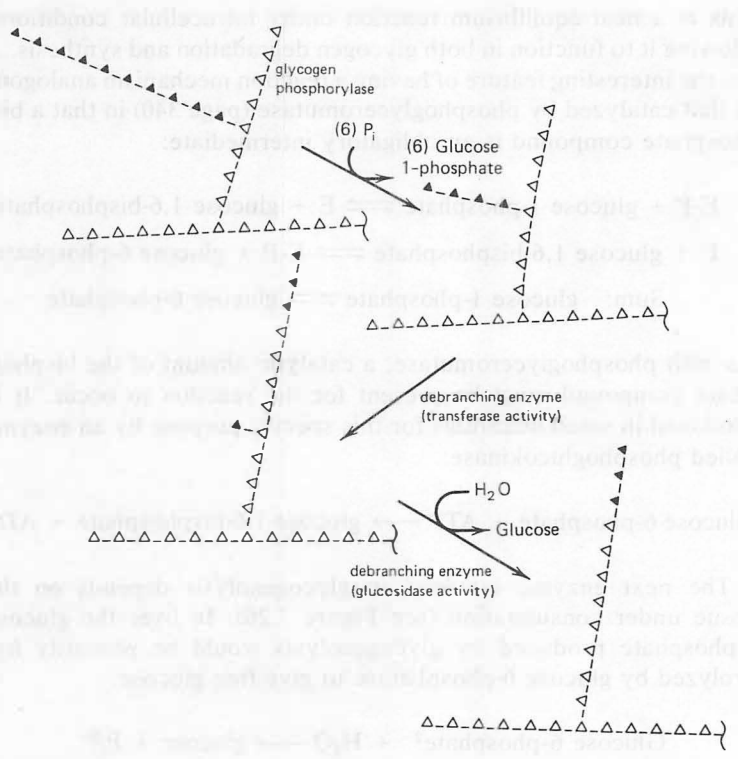


Figure 7.27
Action of the glycogen debranching enzyme.

CLIN. CORR. 7.11 GLYCOGEN STORAGE DISEASES

There are a number of well-characterized glycogen storage diseases, all due to inherited defects of one or more of the enzymes involved in the synthesis and degradation of glycogen. The liver is usually the tissue most affected, but heart and muscle glycogen metabolism can also be defective.

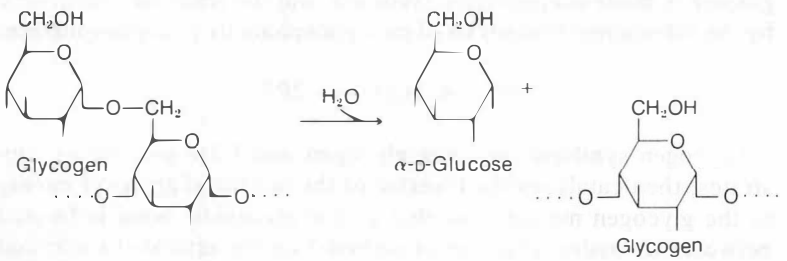
VON GIERKE's DISEASE

The most common glycogen storage disease, referred to as type I or von Gierke's disease, is caused by a deficiency of liver, intestinal mucosa, and kidney glucose 6-phosphatase. Thus diagnosis by small bowel biopsy is possible. This genetic abnormality occurs in only about 1 person in 200,000 and is transmitted as an autosomal recessive trait. Clinical manifestations include fasting hypoglycemia, lactic acidemia, hyperlipidemia, and hyperuricemia with gouty arthritis. The fasting hypoglycemia is readily explained as a consequence of the glucose 6-phosphatase deficiency, the enzyme required to obtain glucose from liver glycogen and gluconeogenesis. The liver of these patients does release some glucose by the action of the glycogen debrancher enzyme. The lactic acidemia occurs because the liver can not use lactate effectively for glucose synthesis. In addition, the liver inappropriately produces lactic acid in response to glucagon. This hormone should trigger glucose release without lactate production; however, the opposite occurs because of the lack of glucose 6-phosphatase. Hyperuricemia results from increased purine degradation in

equation for glycogenolysis followed by glycolysis is

$$(\text{Glucose})_n + 3\text{ADP}^{3-} + 3\text{P}_i^{2-} + \text{H}^+ \longrightarrow (\text{glucose})_{n-1} + 2\ \text{lactate}^{-1} + 3\text{ATP}^{4-}$$

Up to this point in our consideration of glycogenolysis, we have been able to ignore a rather messy feature caused by all the branches that exist in the glycogen molecule. The first enzyme involved in glycogen degradation, glycogen phosphorylase, is specific for α-[1 → 4] glycosidic linkages. It does not even like to go near α-[1 → 6] linkages. Indeed it stops phosphorolyzing α-[1 → 4] glycosidic linkages four glucosyl residues from an α-[1 → 6] branch point. A glycogen molecule that has been degraded to the limit by phosphorylase would look like a well-trimmed hedge, but instead is called the phosphorylase limit dextrin. The action within cells of a "debranching" enzyme is what allows glycogen phosphorylase to continue to degrade glycogen. This "debranching" enzyme is unusual in that it catalyzes two rather different types of reactions, both necessary for the debranching of glycogen. The first is 4-α-D-glucanotransferase activity in which a block of three glucosyl residues is removed from a four-glucosyl residue branch of the glycogen molecule (see Figure 7.27). The block remains covalently attached to the enzyme until it can be transferred to a free 4-hydroxyl of a glucosyl residue at the end of the same or an adjacent glycogen molecule (see Figure 7.27). The result is a longer amylose chain with only one glucosyl residue remaining in [1 6] linkage. This linkage is broken hydrolytically by the other enzyme action of the "debranching" enzyme, that is, its amylo-α-[1,6]-glucosidase activity:



The cooperative and repetitive action of phosphorylase and debranching enzyme results in the complete phosphorolysis and/or hydrolysis of the glycogen molecule. The average molecule of glycogen gives about 12 molecules of glucose 1-phosphate by the action of phosphorylase for every molecule of free glucose produced by the action of the debranching enzyme.

the liver; hyperlipidemia because of increased availability of lactic acid for lipogenesis and chronic lipid mobilization from the adipose tissue.

POMPE'S DISEASE

Type II glycogen storage disease or Pompe's disease is caused by the absence of α-1,4-glucosidase (or acid maltase), an enzyme normally found in lysosomes. The absence of this enzyme leads to the accumulation of glycogen in virtually every tissue. This is somewhat surprising, but lysosomes take up glycogen granules and become defective with respect to other functions if they lack the capacity to destroy the granules. Because other synthetic and degradative pathways of glycogen metabolism are intact, metabolic derangements such as those in von Gierke's disease are not seen. The reason for extralysosomal glycogen accumulation is unknown. Massive cardiomegaly occurs and death results at an early age from heart failure.

CORI'S DISEASE

Also called type III glycogen storage disease, Cori's disease is caused by a deficiency of the glycogen debrancher enzyme. Glycogen accumulates because only the outer branches can be removed from the molecule by phosphorylase. Hepatomegaly occurs, but diminishes with age. The clinical manifestations are similar to but much milder than those seen in von Gierke's disease, because gluconeogenesis is unaffected, and hypoglycemia and its complications are less severe.

McARDLE'S DISEASE

Also called type V glycogen storage disease, McArdle's disease is caused by an

absence of muscle phosphorylase. Patients suffer from painful muscle cramps and are unable to perform strenuous exercise, presumably because muscle glycogen stores are not available to the exercising muscle. Thus, the normal increase in plasma lactate (released from muscle) following exercise is absent. The muscles are probably damaged because of inadequate energy supply and glycogen accumulation. Release of muscle enzymes (creatine phosphokinase, aldolase, myoglobin) is common.

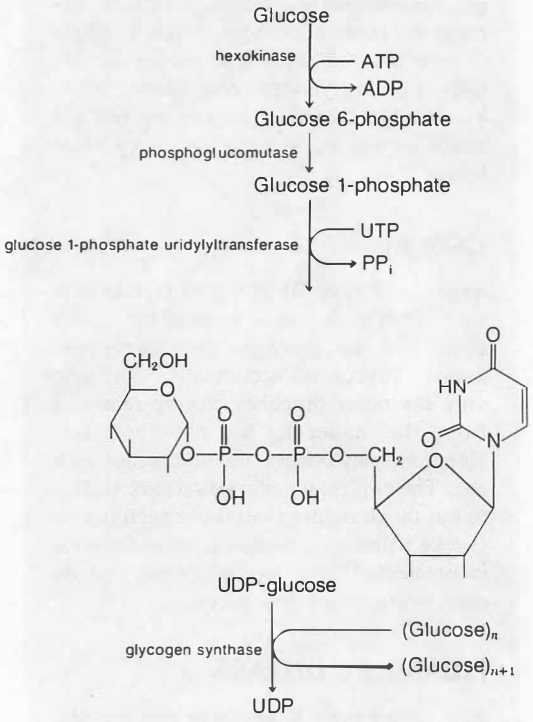


Figure 7.28
Pathway of glycogen synthesis.

There is another minor pathway for glycogen degradation which is quantitatively not that important. Major problems result, however, when this pathway is defective in an individual. As discussed in Clin. Corr. 7.11, a glycosidase of lysosomes degrades glycogen which has been phagocytized, perhaps inadvertently, into these structures during normal turnover of intracellular components.

Pathway of Glycogen Synthesis

The pathway involved in glycogen synthesis is given in Figure 7.28. The first reaction is already familiar, being catalyzed by glucokinase in hepatic tissue and hexokinase in peripheral tissues:

$$\text{Glucose} + \text{ATP} \longrightarrow \text{glucose 6-phosphate} + \text{ADP}$$

The next enzyme involved, phosphoglucomutase, was discussed in relation to glycogen degradation, although this reversible reaction was written in the opposite direction:

$$\text{Glucose 6-phosphate} \longrightarrow \text{glucose 1-phosphate}$$

A unique reaction, finally found in the next step, involves the formation of UDP-glucose by the action of glucose 1-phosphate uridylyltransferase:

$$\text{Glucose 1-phosphate} + \text{UTP} \longrightarrow \text{UDP-glucose} + PP_i$$

This reaction generates an "activated" glucosyl residue, which can be used to build the glycogen molecule. The formation of UDP-glucose is made energetically favorable and the reaction irreversible by the subsequent hydrolysis of pyrophosphate by pyrophosphatase:

$$PP_i^{4-} + H_2O \longrightarrow 2P_i^{2-}$$

Glycogen synthase, utilizing glycogen and UDP-glucose as substrates, then catalyzes the transfer of the activated glucosyl moiety to the glycogen molecule so that a new glycosidic bond is formed between the hydroxyl group of carbon-1 of the activated sugar and carbon-4 of a glucosyl residue of the growing glycogen chain. The reducing end of glucose (carbon-1 of glucose is an aldehyde that can reduce other compounds by donating electrons by the half-reaction of

$$RCH + H_2O \longrightarrow RCOH + 2H^+ + 2e^-)$$

is always added to a nonreducing end of the glycogen chain. Note that the glycogen molecule, regardless of its size, theoretically has only one free reducing end tucked away within the core. It should be noted also that UDP, *not* UMP, is the product of the reaction catalyzed by glycogen synthase. UDP can be converted back to UTP by the action of nucleoside diphosphokinase:

$$\text{UDP} + \text{ATP} \longrightarrow \text{UTP} + \text{ADP}$$

Glycogen synthase is very specific, that is, it will create chains of glucose molecules with α-[1 → 4] linkages but will not participate in the placement of the α-[1 → 6] branches. Its action alone would only produce amylose, the straight-chain polymer of glucose with α-[1 → 4] linkages. Once an amylose chain of at least 10 residues has been formed, a "branching" enzyme comes into play. Its actual name is 1,4-α-glucan branching enzyme because what it does is remove a block of glycosyl residues from a growing chain and transfers it to another chain where it produces an α-[1 6] linkage (see Figure 7.29). The last peculiarity to be mentioned is that the new branch has to be introduced at least four glucosyl residues from an adjacent branch point. Thus the creation of the highly branched structure of glycogen requires the concerted efforts of glycogen synthase and the branching enzyme. The overall balanced equation for glycogen synthesis by the pathway just outlined is

$$(\text{Glucose})_n + \text{glucose} + 2\text{ATP} \longrightarrow (\text{glucose})_{n+1} + 2\text{ADP} + 2\text{P}_i$$

As noted above, the combination of glycogenolysis and glycolysis yields only three molecules of ATP:

$$(\text{Glucose})_n + 3\text{ADP} + 3\text{P}_i \longrightarrow (\text{glucose})_{n-1} + 2 \text{ lactate} + 3\text{ATP}$$

Thus the combination of glycogen synthesis plus glycogen degradation to lactate actually yields the cell only 1 ATP, that is, the sum of the last two equations is

$$\text{Glucose} + \text{ADP} + \text{P}_i \longrightarrow 2 \text{ lactate} + \text{ATP}$$

It should be realized, however, that glycogen synthesis and degradation are carried out during different time frames in a cell. For example, white muscle fibers synthesize glycogen only while at rest when glucose is plentiful and ATP for muscle contraction is not needed. Glycogen is then used during periods of exertion. Although in such terms glycogen storage is not a very efficient process, it provides cells with a fuel reserve that can be quickly and efficiently mobilized.

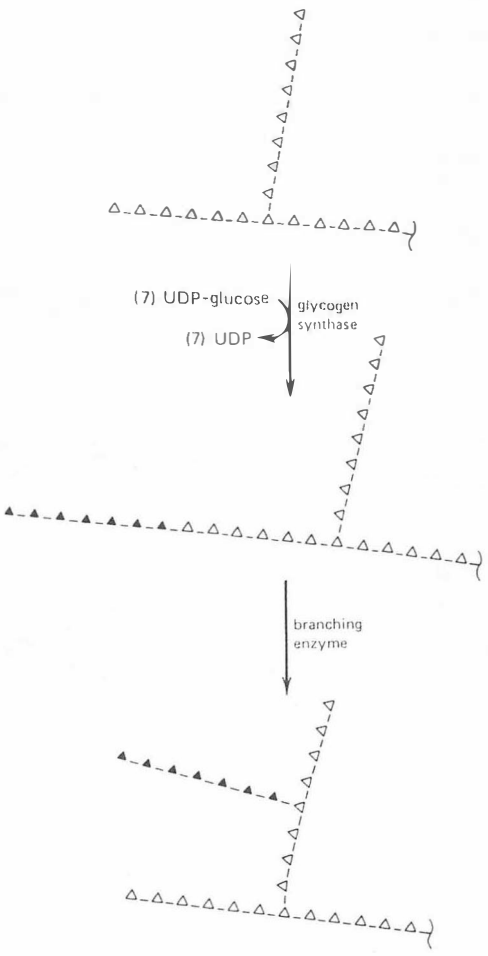


Figure 7.29
Action of the glycogen branching enzyme.

Special Features of Glycogen Degradation and Synthesis

Since glycogen is such a good fuel reserve, it is obvious why we synthesize and store glycogen in liver and muscle. But a good question is why store glucose as glycogen? Why not, for example, store our excess glucose calories entirely as fat instead of glycogen? The answer is at least threefold: (1) we do store fat, some of us lots of it, but fat cannot be mobilized as rapidly in muscle as glycogen; (2) fat cannot be used as a source of energy in the absence of oxygen; and (3) fat cannot be converted to glucose by any pathway of the human body in order to maintain blood glucose levels for use by tissues such as the brain. Why not just pump glucose into cells and store it as free glucose until needed? Why waste so much ATP making a polymer out of glucose, that is, in effect 1 ATP molecule per glucosyl unit? The problem is that glucose is osmotically active. It would cost ATP to "pump" glucose into a cell, regardless of the mechanism, and glucose would have to reach concentrations of 400 mM in liver cells to match the "glucose reserve" provided by the usual liver glycogen levels. Unless balanced by the outward movement of some other osmotically active compound, the accumulation of such concentrations of glucose would cause the uptake of considerable water and the osmotic lysis of the cell. Assuming the molecular mass of a glycogen molecule is of the order of 10^7 daltons, 400 mM glucose is in effect stored at an intracellular glycogen concentration of 0.01 μM. Storage of glucose as glycogen, therefore, creates absolutely no osmotic pressure problem for the cell.

Another interesting feature about glycogen is that a primer is needed for its synthesis. No template is required, but like DNA synthesis, a primer is necessary. Glycogen itself is the usual primer, that is, glycogen synthesis usually takes place by the addition of glucosyl units to glycogen "core" molecules, which are almost invariably present in the cell. The tiers of the glycogen molecule actually get removed and resynthesized much more rapidly than the inner core. Glycogen within a cell is frequently sheared by the combined actions of glycogen phosphorylase and debranching enzyme but is seldom ever obliterated before glycogen synthase and branching enzyme rebuild the molecule. This is a good time to point out why nature has evolved such an elaborate mechanism for creating and disposing of the branched structure of glycogen. In other words, why is glycogen a branched molecule with only one real beginning (the reducing end) and many branches terminating with nonreducing glucosyl units? The answer is that this gives numerous sites of attack for glycogen phosphorylase on a mature glycogen molecule and the same number of sites for glycogen synthase to add an activated

glucosyl unit. If cells synthesized amylose, that is, an unbranched glucose polymer, there would only be one nonreducing end per molecule. The result would be that glycogen degradation and synthesis would surely be much slower processes. As it is, glycogen phosphorylase and glycogen synthase are usually found in tight association with glycogen granules in a cell, as though they exist in the branches of the glycogen tree with ready access to a multitude of nonreducing sugars at the ends of its limbs.

We digressed, however, from the problem of a need of a primer for glycogen synthesis. Perhaps as a consequence of the great number of nonreducing ends, glycogen synthase has a very low K_m for very large glycogen molecules. However, the K_m gets larger and larger as the glycogen molecule gets smaller and smaller. This phenomenon is so pronounced that it is clear that glucose, at its physiological concentration, could never function as a primer. This led to the notion that glycogen must be immortal, that is, that glycogen must be handed down from one cell generation to the next in order for glycogen to be synthesized. Although immortality is attractive, it is now thought that one or more proteins probably function as primers for glycogen synthesis. The hydroxyl groups of seryl or threonyl residues of certain proteins may become glycosylated and then serve as a nucleus for the synthesis of glycogen. Hydrolytic breakage of one of the resulting branches of the glycogen tree rooted in the side of the protein could then provide additional primer for the synthesis of glycogen, independent of any further need for such a protein. Alas, glycogen is probably not immortal.

If glycogen synthase becomes more efficient as the glycogen molecule gets bigger, we ought to worry somewhat about how synthesis of this ball of sugar is curtailed. Fat cells have an almost unlimited capacity to pack away fat—but then they (fat cells) don't have to do anything else. Muscle cells have to participate in mechanical activity and liver cells carry out many processes other than glycogen synthesis. Even in the face of excess glucose, there has to be a way to limit the intracellular accumulation of glycogen. It turns out that glycogen itself inhibits glycogen synthase, the regulatory enzyme involved in glycogen synthesis. Inhibition of the enzyme by glycogen is complicated, and will not be discussed until we have had a chance to present more details with respect to the regulatory enzymes involved in glycogen metabolism.

As stressed in the clinical correlations given in this text, biochemistry has made many significant contributions and is relevant without question to modern clinical medicine. However, in all honesty clinical medicine has contributed and continues to contribute greatly to our understanding of the biochemistry of a number of complex biochemical processes. A case in point is the enzymes ac-

tually involved in glycogen synthesis and degradation. At one time the enzyme glycogen phosphorylase was believed responsible for both the synthesis and degradation of glycogen. This enzyme is responsible for glycogen degradation in the cell but can be readily assayed in the test tube in the direction of glycogen synthesis:

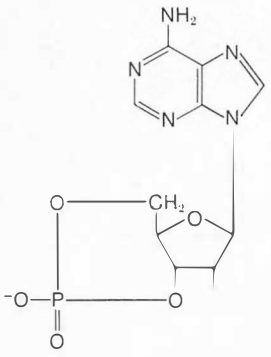
$$(\text{Glucose})_n + \alpha\text{-D-glucose 1-phosphate} \longrightarrow (\text{glucose})_{n+1} + P_i$$

The K'_{eq} for this reaction is approximately unity, and since one of the substrates $(\text{glucose})_n$ looks almost identical to one of the products $(\text{glucose})_{n+1}$ to both us and the enzyme, the direction of flux is determined by the ratio of the steady-state concentrations of P_i and glucose 1-phosphate. Measurements of the concentrations of these components under intracellular conditions make it clear that the reaction is nonequilibrium and that net flux can only be in the direction of net degradation of glycogen. In other words, the intracellular concentration of inorganic phosphate is always much greater than that of glucose 1-phosphate. However, it really did not become clear that glycogen phosphorylase is involved only in glycogen degradation until studies were conducted with a patient who presented with a rare glycogen storage disease, now known as McArdle's disease. In this type of glycogenosis, the enzyme glycogen phosphorylase of skeletal muscle is missing (see Clin. Corr. 7.11). Nevertheless, the skeletal muscles of patients with McArdle's disease are loaded with glycogen. This observation made it obvious that another enzyme had to be involved in the synthesis of glycogen, and helped lead to the isolation and characterization of glycogen synthase.

Regulation of Glycogen Synthesis and Degradation

Regulators of the Pathways

Glycogen synthase and glycogen phosphorylase are the regulatory enzymes of glycogen synthesis and degradation, respectively. Both catalyze nonequilibrium reactions, both are subject to control by effectors, and both are subject to covalent modulation. Figure 7.30 presents—without the frills—an overview of the regulation of glycogen metabolism. Effector control by metabolites is believed to be physiologically important. Glucose 6-phosphate activates glycogen synthase, whereas AMP activates glycogen phosphorylase. Superimposed upon effector control is a much more complex control system whereby glucagon in the liver and epinephrine in the muscle bring about rapid mobilization of liver and muscle glycogen, respectively. This is accomplished by way of cAMP (see Figure 7.30*b*). As discussed in more detail in Chapter 16, cAMP functions as a



Cyclic AMP

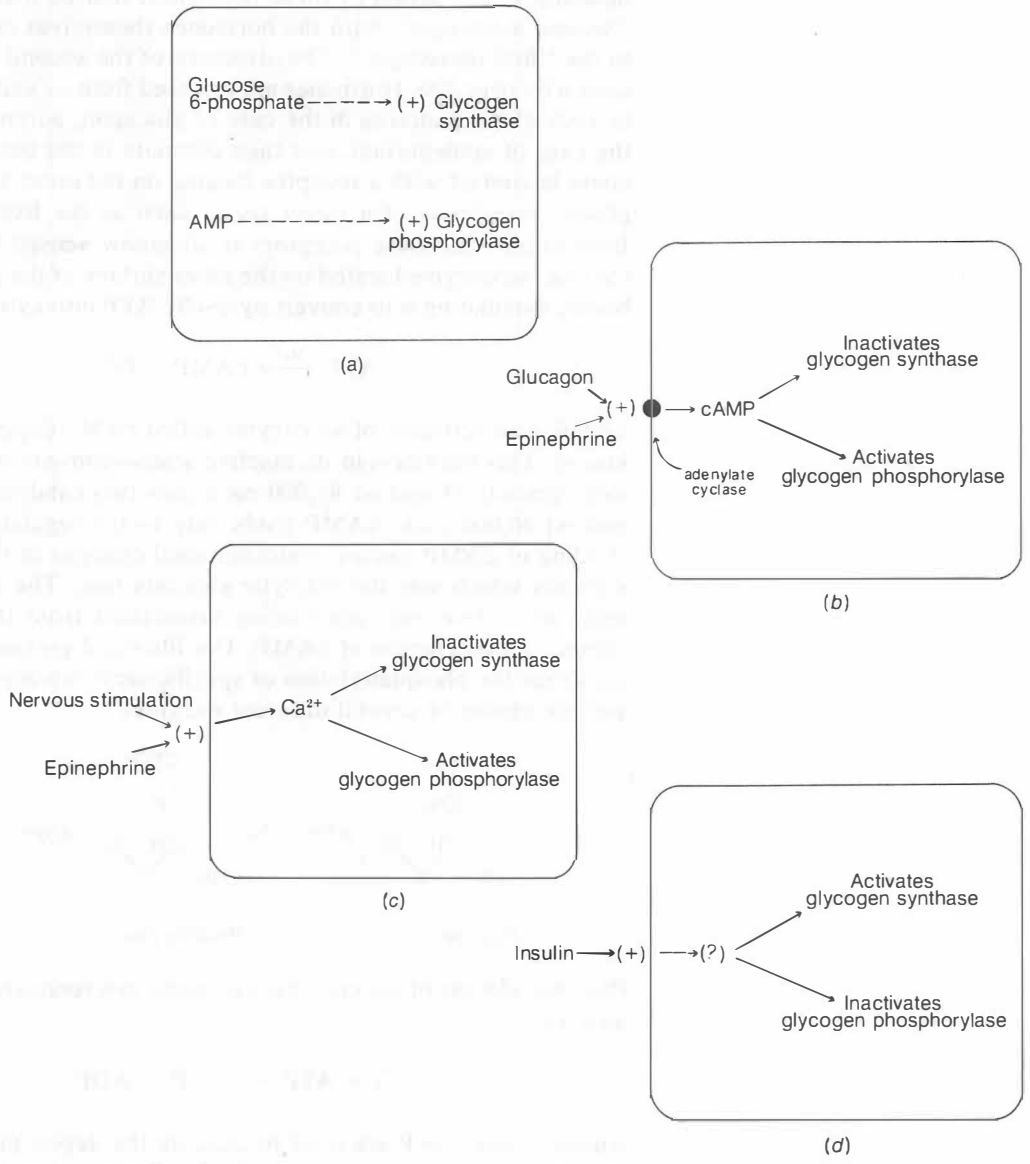
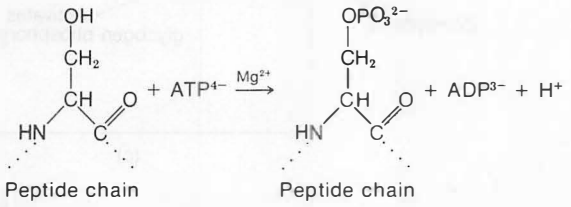


Figure 7.30
Overviews of the mechanisms involved in the regulation of glycogen metabolism.
(a) Effector control. (b) Cyclic AMP promotes glycogen degradation. (c) Ca^{2+} promotes glycogen degradation. (d) Insulin promotes glycogen synthesis.

mediator of the action of these hormones, that is, it functions as a "second messenger" with the hormones themselves corresponding to the "first messenger." The structure of the second messenger is shown on page 390. Hormones are released from an endocrine tissue (α cells of the pancreas in the case of glucagon; adrenal medulla in the case of epinephrine), and then circulate in the blood until they come in contact with a receptor located on the outer surface of the plasma membrane of a target tissue such as the liver or muscle. Interaction with these receptors is somehow sensed by adenylate cyclase, an enzyme located on the inner surface of the plasma membrane, stimulating it to convert cytosolic ATP into cytosolic cAMP:

$$\text{ATP} \xrightarrow{Mg^{2+}} \text{cAMP} + PP_i$$

cAMP is an activator of an enzyme called cAMP-dependent protein kinase. This enzyme—in its inactive state—consists of two regulatory subunits of mol wt 85,000 each plus two catalytic subunits of mol wt 40,000 each. cAMP binds only to the regulatory subunits. Binding of cAMP causes conformational changes in the regulatory subunits which sets the catalytic subunits free. The catalytic subunits are active only after being dissociated from the regulatory subunit by this action of cAMP. The liberated protein kinase then catalyzes the phosphorylation of specific seryl residues of the polypeptide chains of several different enzymes.



Phosphorylation of an enzyme can more conveniently be abbreviated as

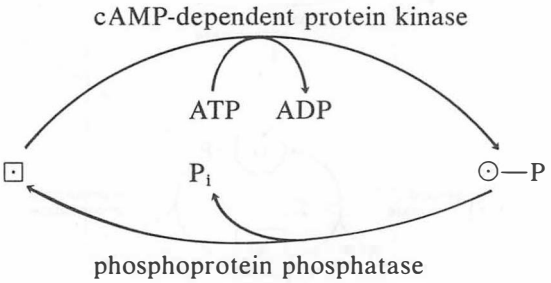
$$\boxdot + \text{ATP} \rightarrow \odot\text{—P} + \text{ADP}$$

where $\boxdot$ and $\odot$—P are used to indicate the dephosphorylated and phosphorylated enzymes, respectively. The circle and square symbols are used because phosphorylation of enzymes subject to regulation by covalent modulation causes a change in their conformation, which affects the active site. It turns out that the change in conformation due to phosphorylation greatly increases the catalytic activity of some enzymes but greatly decreases the catalytic activity of others. It depends on the enzyme involved. Only a few of the known

enzymes are subject to this type of regulation, called either covalent modulation or covalent modification. Regardless of whether phosphorylation or dephosphorylation activates the enzyme, the active form of the enzyme is called the "a" form and the inactive form the "b" form. Likewise, regardless of the effect of phosphorylation on catalytic activity, the action of cAMP-dependent protein kinase is always opposed by a phosphoprotein phosphatase, which catalyzes the reaction of

$$\odot\text{—P} + H_2O \longrightarrow \boxdot + P_i$$

Putting these together creates a cyclic control system, such that the ratio of phosphorylated enzyme to dephosphorylated enzyme is a function of the relative activities of the cAMP-dependent protein kinase and the phosphoprotein phosphatase.



If the kinase has greater activity than the phosphatase, more enzyme will be in the phosphorylated mode—and vice versa. Since the activity of an interconvertible enzyme (i.e., an enzyme subject to covalent modulation) is determined by whether it is in the phosphorylated or dephosphorylated mode, the relative activities of the kinase and phosphatase dictate the portion of a particular enzyme which is in the catalytically active state.

Calcium (Ca^{2+}) can also function as a "second messenger" (see Figure 7.30*c*). It activates a different kinase, completely distinct from cAMP-dependent protein kinase, which has been named phosphorylase kinase. It picked up this name because, as is discussed below, it is responsible for the phosphorylation (and activation) of glycogen phosphorylase. However, we now know that this kinase will phosphorylate a number of other enzymes as well, and perhaps could be appropriately renamed Ca^{2+}-dependent protein kinase. Regardless of what you call it, phosphorylase kinase is an interesting enzyme, consisting of four nonidentical subunits with an overall molecular weight of 1.3×10^6. The subunit that confers Ca^{2+} sensitivity upon the enzyme has been named calmodulin—meaning Ca^{2+}-

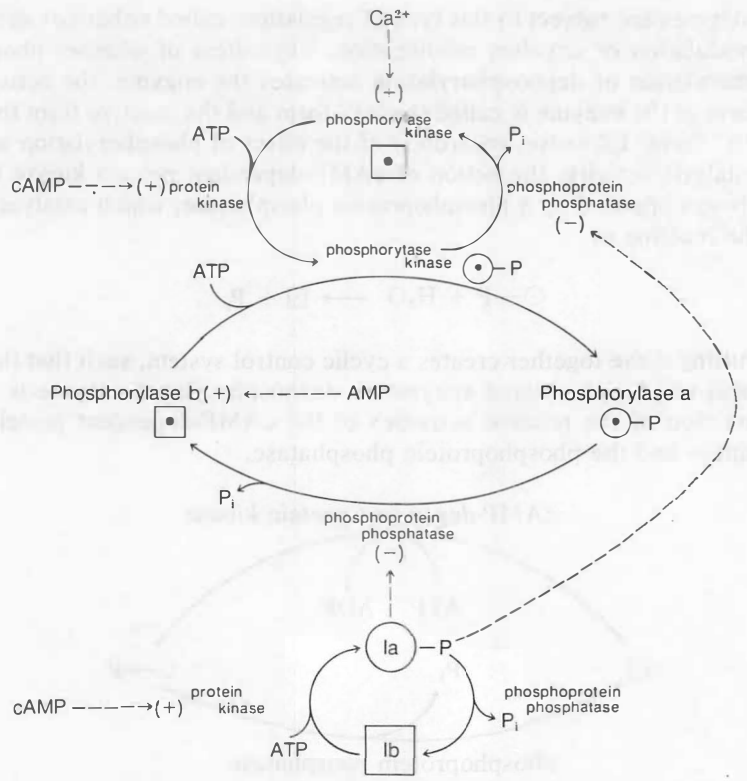
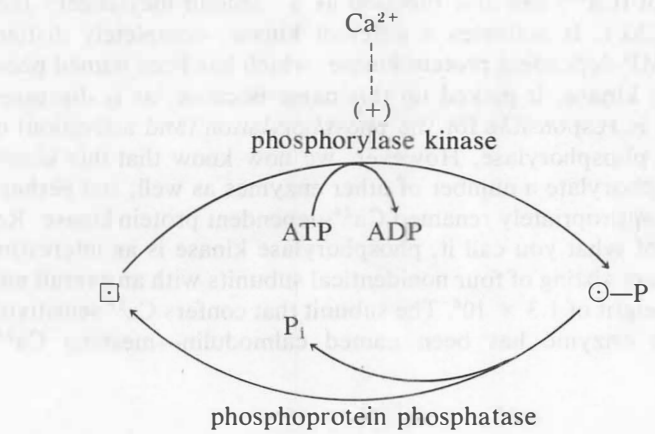


Figure 7.31
Regulation of glycogen phosphorylase.
Ia refers to the a or phosphorylated form of the inhibitory protein; Ib to the b or non-phosphorylated form.

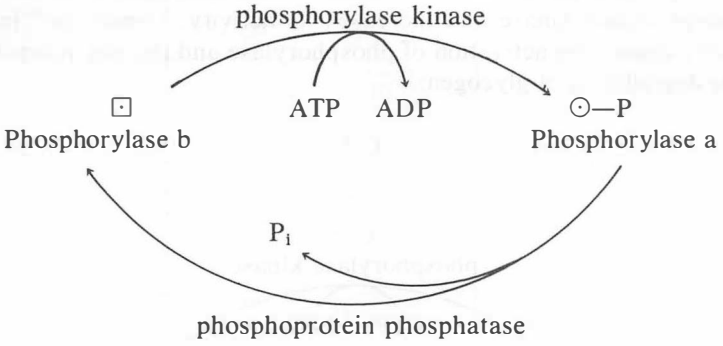
modulating protein. Calmodulin is by no means unique to phosphorylase kinase, but rather is responsible for the Ca^{2+} sensitivity of several other enzyme complexes as well. The cyclic process by which Ca^{2+} can modulate the activity of enzymes subject to covalent modulation can be abbreviated as follows:



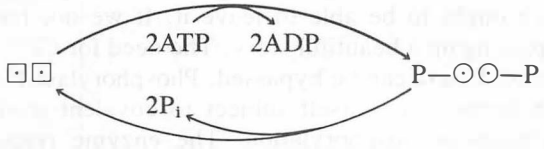
Now with this background information, we can proceed with a discussion of how glycogen metabolism is regulated.

Regulation of Glycogen Phosphorylase

The mechanisms responsible for the regulation of glycogen phosphorylase are summarized in a rather formidable fashion in Figure 7.31. The enzyme is subject to allosteric inhibition by glucose 6-phosphate and ATP and allosteric activation by P_i, AMP, and IMP. These effectors are considered of some physiological significance in the regulation of glycogen metabolism, with the activating effects of nucleotide monophosphates being the ones to remember. Nevertheless, effector control of phosphorylase has to be considered "primitive" with respect to its very elaborate control by covalent modulation. Phosphorylase exists in an "a" form, which is active, and a "b" form, which is inactive. These forms of the enzyme are interconverted by the actions of phosphorylase kinase and phosphoprotein phosphatase:



Although there is still some uncertainty, phosphorylase appears to be composed under physiological conditions of two identical, 92,000 mol wt subunits. Serine-14, counting from the amino terminal of the subunits becomes phosphorylated as a result of the action of phosphorylase kinase. Hence, the last reaction should really be written as

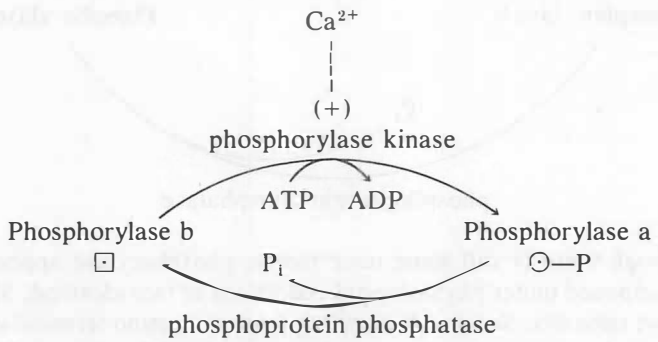


However, we will be less formal on this point, and will usually ignore how many subunits are involved, and also how many times each subunit becomes phosphorylated.

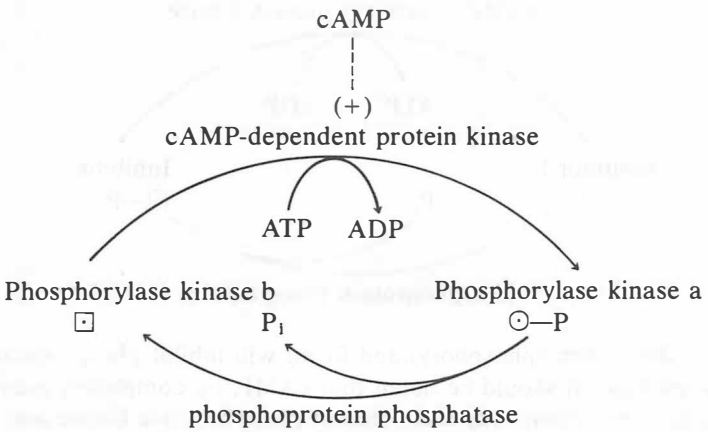
A conformational change caused by phosphorylation of phos-

phorylase b transforms the enzyme into a catalytically active state. Independently of the phosphorylation mechanism, AMP and P_i can greatly activate phosphorylase b. These effectors have little effect, however, on the already quite active phosphorylase a. Hence the "primitive" allosteric mechanism is bypassed by the covalent modulation mechanism. One can, of course, look at it the other way, that is, the covalent modulation mechanism can be bypassed by the "primitive" allosteric mechanism. Although AMP does not further activate phosphorylase a, it does bind to phosphorylase and make this enzyme a poor substrate for the action of phosphoprotein phosphatase. Thus, AMP not only activates phosphorylase b but also protects phosphorylase a from being inactivated by dephosphorylation. The interaction of an effector (AMP) with a substrate (phosphorylase a) in such a way as to influence the activity of an enzyme (phosphoprotein phosphatase) may seem rather unusual. However, it is a very common phenomenon with enzymes subject to covalent modulation, and is believed to be of considerable physiological significance.

As pointed out above, Ca^{2+} binds to the calmodulin subunit of phosphorylase kinase and increases its activity. Hence Ca^{2+} indirectly causes the activation of phosphorylase and thereby promotes the degradation of glycogen.



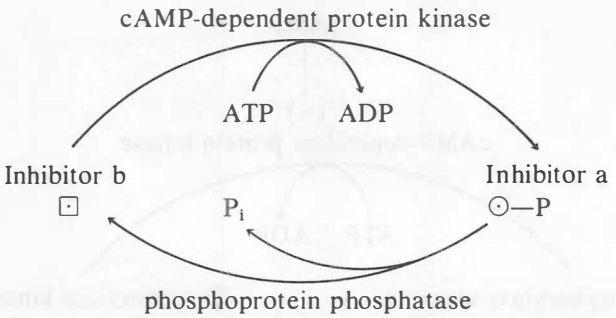
It would seem that this is surely enough control on phosphorylase—and that we ought to be able to leave it. If we do, however, we would be passing up a beautiful story. The need for Ca^{2+} to activate phosphorylase kinase can be bypassed. Phosphorylase kinase exists in a and b forms and is itself subject to covalent modulation by phosphorylation–dephosphorylation. The enzyme responsible for its phosphorylation is the cAMP-dependent protein kinase.



As with phosphorylase, the phosphorylated form of the enzyme is the active species.

The cAMP-dependent protein kinase cannot phosphorylate phosphorylase directly—it has to exert its effects via phosphorylase kinase. Now if you combine these two cyclic systems, as has been done in Figure 7.31, it can be seen that phosphorylase is subject to regulation by a bicyclic system.

It is obvious that regulation of phosphorylase kinase will have a substantial effect on the activity of glycogen phosphorylase. It is equally obvious that turning on and off phosphoprotein phosphatase could achieve the same thing. But what would really provide ultimate control for the activation of phosphorylase would involve the simultaneous turning off of phosphoprotein phosphatase and turning on of phosphorylase kinase, and vice versa, for the inactivation of the enzyme. Since phosphoprotein phosphatase also acts on phosphorylase kinase, turning off phosphoprotein phosphatase would also achieve greater activation of phosphorylase kinase. It appears from recent work that such a mechanism exists for this sort of reciprocal relationship, with just a slight twist to make it interesting. A protein exists in cells which inhibits phosphoprotein phosphatase. This protein is called inhibitory protein (I), but it also is subject to covalent modulation by phosphorylation–dephosphorylation by cAMP-dependent protein kinase and phosphoprotein phosphatase, respectively:



Only the a form (phosphorylated form) will inhibit phosphoprotein phosphatase. It should be noted that cAMP, by completely indirect mechanisms, causes the activation of phosphorylase kinase and the inhibition of phosphoprotein phosphatase—making it possible theoretically to completely activate glycogen phosphorylase (see Figure 7.31).

It is interesting that the phosphorylated inhibitor is converted back to its dephosphorylated form by the enzyme it inhibits—the phosphoprotein phosphatase! However, the phosphorylated inhibitor cleverly does not inhibit its own dephosphorylation, just the dephosphorylation of phosphorylase kinase a and phosphorylase a. In contrast to all other interconvertible enzymes that become phosphorylated on serine residues, inhibitor protein becomes phosphorylated on a threonine residue. This may account in part for why it can inhibit the action of phosphoprotein phosphatase against other phosphorylated enzymes and yet serve itself as a substrate for phosphoprotein phosphatase.

Now if you have not already given up because of the complexities of the glycogen phosphorylase regulatory system, note that there is a good reason for the existence of the bicyclic control system for the phosphorylation of phosphorylase plus the additional control on its dephosphorylation. This provides a tremendous amplification mechanism. Think about it with relation to Figure 7.31. One molecule of epinephrine or one molecule of glucagon can cause, by the activation of adenylate cyclase, the formation of many molecules of cAMP. cAMP can then activate cAMP-dependent protein kinase, which, in turn, can cause the activation of many molecules of phosphoprotein phosphatase inhibitor as well as the activation of many molecules of phosphorylase kinase. In turn, phosphorylase kinase can cause the phosphorylation of many molecules of glycogen phosphorylase—which in turn can cause the phosphorolysis of many glycosidic bonds of glycogen. This entire system is an elaborate amplification system in which the signal provided by just a few molecules of hormone can be amplified into production of an enormous number of

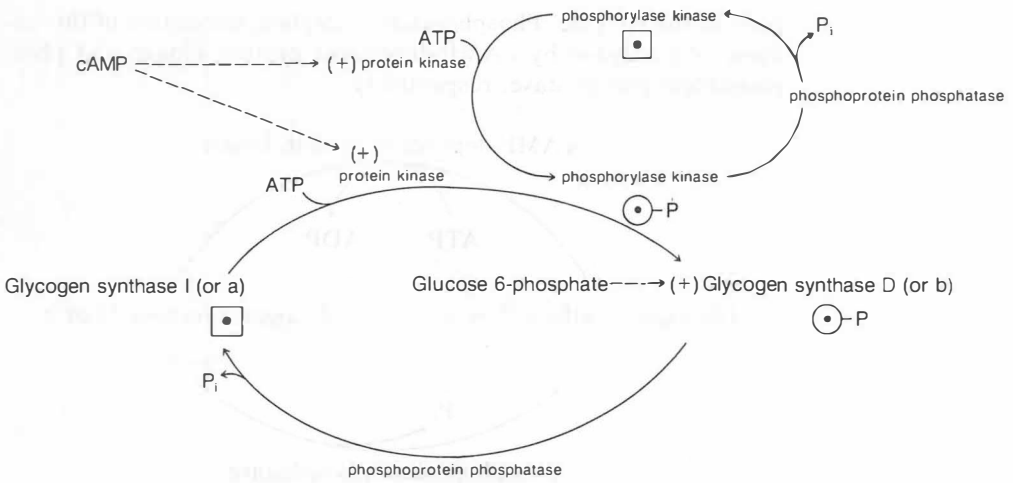


Figure 7.32
Regulation of glycogen synthase.

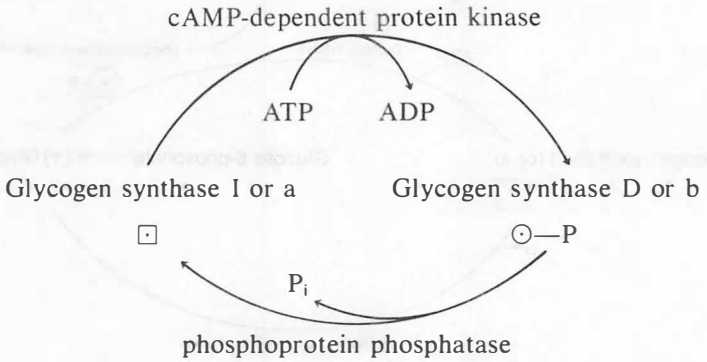
glucose 1-phosphate molecules. If each step represents an amplification factor of 100, then a total of 5 steps would result in an amplification of 10 billion! This system is so rapid, in large part because of the amplification system, that all of the stored glycogen of white muscle fibers could be completely mobilized within just a few seconds.

Regulation of Glycogen Synthase

Glycogen synthase is the regulatory enzyme involved in glycogen synthesis. It has to be active for glycogen synthesis and inactive during glycogen degradation. The combination of the reactions catalyzed by glycogen synthase, glycogen phosphorylase, glucose 1-phosphate uridylyltransferase, and nucleoside diphosphokinase adds up to a futile cycle with the overall equation: ATP → ADP + P_i. Hence glycogen synthase needs to be turned off when glycogen phosphorylase is turned on, and vice versa.

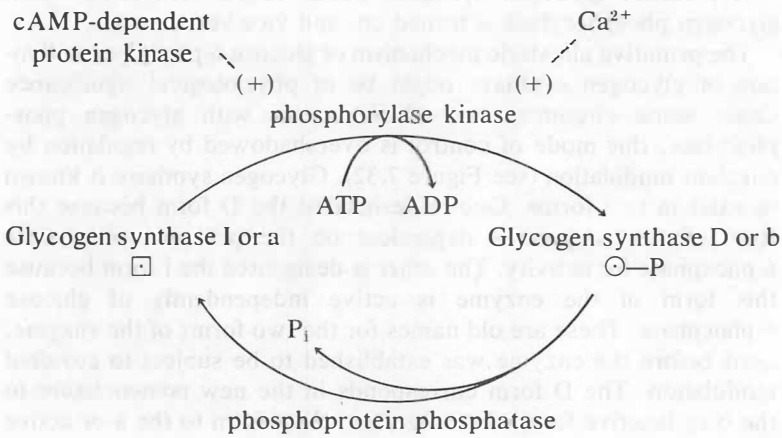
The primitive allosteric mechanism of glucose 6-phosphate activation of glycogen synthase might be of physiological significance under some circumstances. However, as with glycogen phosphorylase, this mode of control is overshadowed by regulation by covalent modulation (see Figure 7.32). Glycogen synthase is known to exist in two forms. One is designated the D form because this form of the enzyme is dependent on the presence of glycose 6-phosphate for activity. The other is designated the I form because this form of the enzyme is active independently of glucose 6-phosphate. These are old names for the two forms of the enzyme, used before the enzyme was established to be subject to covalent modulation. The D form corresponds in the new nomenclature to the b or inactive form of the enzyme, the I form to the a or active

form of the enzyme. Phosphorylation–dephosphorylation of this enzyme is catalyzed by cAMP-dependent protein kinase and phosphoprotein phosphatase, respectively,



Note that in contrast to phosphorylase, phosphorylation of glycogen synthase causes inactivation of the enzyme. It should be immediately appreciated that cAMP is an intracellular signal for reciprocally controlling glycogen synthase and glycogen phosphorylase. In contrast to the regulation of glycogen phosphorylase, regulation of glycogen synthase by phosphorylation is not bicyclic. In this case cAMP-dependent protein kinase can directly phosphorylate glycogen synthase, bypassing the need for phosphorylase kinase. Not shown in Figure 7.32, cAMP probably exerts an additional effect on glycogen synthase by indirect effects on phosphoprotein phosphatase via the phosphorylated inhibitor.

Phosphorylase kinase is also able to phosphorylate and inactivate glycogen synthase:



This is important because phosphorylase kinase, in contrast to cAMP-dependent protein kinase, is sensitive to regulation by Ca^{2+}. Thus, both cAMP and Ca^{2+} are able to regulate the two regulatory enzymes of glycogen metabolism in a reciprocal manner. Exactly why cAMP-dependent protein kinase directly phosphorylates glycogen synthase, thereby eliminating the need for bicyclic control of glycogen synthase and losing some of the amplification factor, is not known.

Effector Control of Glycogen Metabolism

Referred to above as a "primitive" regulatory mechanism, effector control is clearly important under some physiological conditions. For example, certain muscles under anaerobic conditions have been shown to rapidly mobilize their glycogen stores without marked conversion of phosphorylase b into phosphorylase a nor glycogen synthase a into glycogen synthase b. Presumably this is accomplished by effector control in which ATP levels decrease, causing less inhibition of phosphorylase; glucose 6-phosphate levels decrease, causing less inhibition of phosphorylase and less activation of glycogen synthase; AMP and IMP levels increase, causing activation of phosphorylase; and P_i levels increase, causing activation of phosphorylase. This enables the muscle to keep working, for at least a short period of time, by using the ATP produced by glycolysis of the glucose 6-phosphate obtained from glycogen.

Proof that effector control can operate has also been obtained in studies of a genetic strain of mice which are deficient in muscle phosphorylase kinase. Phosphorylase b in the muscle of such mice cannot be converted into phosphorylase a. Nevertheless, heavy exercise of these mice results in depletion of muscle glycogen, presumably because of stimulation of phosphorylase by effectors.

As mentioned previously, glycogen is able to exert a feedback control over its own formation. This is indirect, accomplished by glycogen inhibition of the dephosphorylation of glycogen synthase b by phosphoprotein phosphatase. This shifts the steady state in favor of glycogen synthase b and limits further glycogen synthesis.

Another important regulation appears to occur specifically in liver. High concentrations of glucose in the liver promote glycogen synthesis by inhibiting glycogen phosphorylase. This is accomplished by two mechanisms: (1) glucose binds to phosphorylase a and inhibits its activity, and (2) the complex of glucose and phosphorylase a is a much better substrate for phosphoprotein phosphatase action than phosphorylase a without glucose. Phosphorylase seems to act as a "glucose receptor" in the liver, increasing in activity when blood glucose is low and decreasing in activity when blood glucose is high.

Mobilization of Tissue Glycogen

The mechanism responsible for glucagon mobilization of liver glycogen is given diagrammatically in Figure 7.30. The cascade initiated by glucagon results in inactivation of glycogen synthase by the mechanism given in Figure 7.32 and activation of glycogen phosphorylase by the mechanism given in Figure 7.31. The net result is a very rapid increase in blood glucose levels.

The mechanism responsible for epinephrine stimulation of muscle glycogen degradation has been described in Figures 7.31 and 7.32. cAMP, produced in response to epinephrine stimulation of adenylate cyclase, signals the concurrent inactivation of glycogen synthase and the activation of glycogen phosphorylase. This makes more substrate (glucose 6-phosphate) available for ATP production by glycolysis.

Epinephrine mobilizes liver glycogen by at least three different mechanisms, with the physiologically most important mechanism not being clearly established for humans. One mechanism involves epinephrine stimulation of glucagon release from the α cells of the pancreas. Glucagon then travels by way of the blood to mobilize liver glycogen as discussed above. Epinephrine can also interact directly with receptors in the plasma membrane of the liver cells to activate adenylate cyclase. The resulting increase in cAMP has the same effect as that caused by glucagon. The binding site for epinephrine on the plasma membrane, which is in communication with adenylase cyclase, is called the β receptor. Although there are considerable species differences and the picture is a bit hazy for the human at this time, the plasma membrane of liver cells also has another binding site for epinephrine, called α receptors. Interaction of epinephrine with these receptors leads to Ca^{2+} release into the cytosol, rather than adenylate cyclase activation. As outlined in Figures 7.30c, 7.31, and 7.32, the increase in Ca^{2+} activates phosphorylase kinase, which in turn activates glycogen phosphorylase and inactivates glycogen synthase. The consequence of all of these mechanisms is the same—increased release of glucose into the blood from the glycogen stored in the liver.

Nervous excitation of muscle activity is mediated via changes in intracellular Ca^{2+} concentrations. The nerve impulse causes membrane depolarization which, in turn, causes Ca^{2+} release from the sarcoplasmic reticulum into the sarcoplasm of muscle cells. This release of Ca^{2+} triggers muscle contraction, whereas reaccumulation of Ca^{2+} by the sarcoplasmic reticulum causes relaxation. The same range of Ca^{2+} concentrations effective in causing muscle contraction (10^{-8}–10^{-6} M) also greatly affects the activity of phosphorylase kinase. As Ca^{2+} concentrations increase there is more muscle activity and a greater need for ATP. The activation of phos-

phorylase kinase by Ca^{2+} leads to the subsequent activation of glycogen phosphorylase and inactivation of glycogen synthase. The result is that more glycogen is converted to glucose 6-phosphate so that more ATP can be produced to meet the greater energy demand of muscle contraction.

Promotion of Glycogen Synthesis in Muscle and Liver by Insulin

Stimulation of glucose uptake or transport across the plasma membrane is one factor involved in the promotion of muscle glycogenesis by insulin. Although the mechanism involved is not clear, insulin is also able to activate glycogen synthase and inactivate glycogen phosphorylase. One current idea is that insulin interacts with a receptor which triggers the production of some intracellular messenger, which, in turn, influences the relative activities of either the protein kinases, the phosphoprotein phosphatases, or both (see Figure 7.30). One proposal is that this messenger influences the equilibrium of the reaction by which cAMP activates protein kinase:

$$R{-}C + cAMP \rightleftharpoons R{-}cAMP + C$$

According to this hypothesis, the second messenger of insulin action shifts the equilibrium of the reaction to the left, making cAMP less effective as an activator of protein kinase.

Little is known as to how insulin activates glycogen synthesis in the liver. It clearly is not related to glucose transport, the liver having a high capacity, insulin-insensitive transport system for glucose. Presumably it relates to the regulation of the interconversion of glycogen synthase and phosphorylase by some intracellular messenger of insulin action. At this time the mechanism responsible for insulin's action upon cells remains one of the major unsolved problems of biochemistry.

BIBLIOGRAPHY

Brooke, M. H., and Kaiser, K. K. The use and abuse of muscle histochemistry. *Ann. N.Y. Acad. Sci.* 228:121, 1974.

Cohen, P. *Control of enzyme activity*. New York: Wiley, 1976.

Denton, R. M., and Pogson, C. I. *Metabolic regulation*. New York: Wiley, 1976.

Geelen, M. J. H., Harris, R. A., Beynen, A. C., and McCune, S. A. Short-term hormonal control of hepatic lipogenesis. *Diabetes* 29:1006, 1980.

Greene, H. L., Slonin, A. E., and Burr, I. M. Type I glycogen storage disease: a metabolic basis for advances in treatment, in L. A. Barness (ed.), *Advances in pediatrics*, vol. 26, Chicago: Year Book Publishers, 1979, p. 63.

Hanson, R. W., Mehlman, M. A. (eds.). *Gluconeogenesis, its regulation in mammalian species.* New York: Wiley, 1976.

Isselbacher, K. J., Adams, R. D., Braundwald, E., Petersdorf, R. B., and Wilson, J. D. (eds.). *Harrison's principles of internal medicine*, 9th ed. New York: McGraw-Hill, 1980.

Klachko, D. M., Anderson, R. R., and Heimberg, M. (eds.). *Hormones and energy metabolism.* New York: Plenum, 1979.

Larner, J. *Intermediary metabolism and its regulation.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Lieber, C. S. The metabolism of alcohol. *Sci. Am.* 234(3):25, 1976.

Mitchell, G., Heffron, J. J. A., and van Rensburg, A. J. J. A halothane-induced biochemical defect in muscle of normal and malignant hyperthermia-susceptible landrace pigs. *Anesth. Analg.* 59:250, 1980.

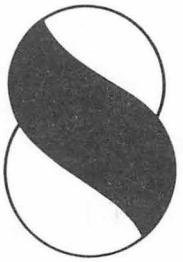
Metzler, D. E. *Biochemistry. the chemical reactions of living cells,* New York, Academic Press, 1977.

Newsholme, E. A., and Start, C. *Regulation in metabolism.* New York: Wiley, 1973.

Roach, P. J. Principles of the regulation of enzyme activity, in L. Goldstein and D. M. Prescott (eds.), *Cell biology, a comprehensive treatise,* vol. IV. New York: Academic, 1980.

Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. (eds.). *The metabolic basis of inherited disease*, 4th ed. New York: McGraw-Hill, 1978.

Stanley, C. A., Anday, E. K., Baker, L., and Delivoria-Papadopolous, M. Metabolic fuel and hormone responses to fasting in newborn infants. *Pediatrics* 64:613, 1979.



Carbohydrate Metabolism II

NANCY B. SCHWARTZ

SPECIAL PATHWAYS

In addition to the catabolism of glucose for the specific purpose of energy production in the form of ATP, several other pathways involving sugar metabolism exist in cells. One alternate pathway of glucose metabolism, the pentose phosphate pathway, is particularly important in animal cells. As will be discussed, the pentose phosphate pathway does not operate instead of glycolysis and the TCA cycle, but rather functions side by side for production of reducing power and pentose intermediates.

Additionally, there are specific pathways for synthesis and degradation of monosaccharides, oligosaccharides, and polysaccharides (other than glycogen), as well as a profusion of chemical interconversions, whereby one sugar can be changed into another. Most of these transformation reactions and the synthesis of oligo- and polysaccharides occur at the sugar nucleotide level.

8.1 PENTOSE PHOSPHATE PATHWAY

Many cells possess another pathway for degradation of glucose to carbon dioxide, known as the pentose phosphate pathway, the hexose monophosphate shunt, or the 6-phosphogluconate pathway. The metabolic significance of this pathway is not to obtain energy from the oxidation of glucose in animal tissues. In fact, starting with glucose 6-phosphate, no ATP is generated, nor is any required. The pentose phosphate pathway is rather a multifunctional pathway whose primary purpose is to generate reducing power in the form of NADPH. It has previously been mentioned that the fundamental distinction between NADH and NADPH in most biochemical reactions is that NADH is oxidized by the respiratory chain to produce ATP, whereas NADPH serves as a hydrogen and electron donor in reductive biosynthetic reactions. It should also be emphasized that the enzymes involved in this pathway are all located in the cytosol, indicating that oxidation is not dependent on the mitochondria or the TCA cycle.

Another important function of this pathway is to convert hexoses into pentoses, particularly ribose 5-phosphate. This five-carbon sugar and its derivatives are components of ATP, CoA, NAD, FAD, RNA, and DNA. The pentose phosphate pathway also catalyzes the interconversion of 3-, 4-, 6-, and 7-carbon sugars, some of which can enter the glycolytic sequence. In order to fulfill another function, which will not be discussed further, the pentose phosphate pathway

may be modified to participate in the formation of glucose from CO_2 in photosynthesis.

Reaction Sequence of Pentose Phosphate Pathway

The oxidative pentose phosphate pathway provides a means for cutting the carbon chain of a sugar molecule one carbon at a time. However, in contrast to glycolysis and the TCA cycle, the operation of this pathway does not occur as a consecutive set of reactions leading directly from glucose 6-phosphate to six molecules of CO_2. For simplification, the overall pathway can be visualized as occurring in two stages. In the first, hexose is decarboxylated to pentose. All of the oxidation reactions that lead to formation of NADPH also occur in this stage. The pathway may continue further, and, by a series of transformations, six molecules of pentose may undergo rearrangements to yield five molecules of hexose. It is these various transformations that give the pentose phosphate pathway its characteristic complexity. To understand this pathway, it is necessary to examine each reaction individually.

Formation of Pentose Phosphate

The first reaction of the pentose phosphate pathway (Figure 8.1) is the enzymatic dehydrogenation of glucose 6-phosphate at C-1 to form 6-phosphogluconate and NADPH. The enzyme catalyzing this reaction is glucose 6-phosphate dehydrogenase, the first enzyme found to be specific for $NADP^+$. Special interest in this enzyme stems from the severe anemia that may result from the absence or presence of glucose 6-phosphate dehydrogenase in erythrocytes of one of several genetic variants (Clin. Corr. 8.1). Formation of the intermediate product of this reaction, a lactone, is freely reversible. Although the lactone is unstable and hydrolyzes spontaneously, a specific gluconolactonase causes a more rapid ring opening and ensures that the reaction goes to completion. The overall equilibrium of these two reactions lies far in the direction of NADPH maintaining a high [NADPH]/[$NADP^+$] ratio within cells. A second dehydrogenation and decarboxylation is catalyzed by 6-phosphogluconate dehydrogenase, a Mg^{2+}-dependent enzyme. The pentose phosphate, ribulose 5-phosphate, and a second molecule of NADPH are produced. The final step in synthesis of ribose 5-phosphate is the isomerization of ribulose 5-phosphate by phosphopentose isomerase. Like similar reactions in the glycolytic pathway, this ketose-aldose isomerization proceeds through an enediol intermediate.

CLIN. CORR. 8.1 GLUCOSE 6-PHOSPHATE DEHYDROGENASE: GENETIC ABSENCE OR PRESENCE OF GENETIC VARIANTS IN ERYTHROCYTES

When certain seemingly harmless drugs, such as antimalarials, are administered to susceptible patients, an acute hemolytic anemia may result. Susceptibility to drug-induced hemolytic disease may be due to a deficiency of glucose 6-phosphate dehydrogenase activity in the erythrocyte, and was one of the early indications that genetic deficiencies of this enzyme exist. This enzyme, which catalyzes the coupled oxidation of glucose 6-phosphate to 6-phosphogluconate and the reduction of NADP, is particularly important, since the pentose phosphate pathway is the only known pathway of oxygen consumption and CO_2 production in the red cell. For example, red cells with the relatively mild A-type of glucose 6-phosphate dehydrogenase deficiency are able to oxidize glucose at a normal rate when the demand for NADPH is normal. However, if the rate of NADPH oxidation is increased, the enzyme-deficient cells are unable to adequately increase the rate of glucose oxidation and carbon dioxide production. Additionally, cells lacking glucose 6-phosphate dehydrogenase do not reduce enough NADP to maintain glutathione in its reduced state. Reduced glutathione is necessary for the integrity of the erythrocyte membrane, thus rendering enzyme-deficient red cells more susceptible to hemolysis by a wide range of compounds. Therefore, the basic abnormality in glucose 6-phosphate deficiency is the formation of mature red blood cells that have

These first reactions resulting in decarboxylation may be considered to be the most important, since all the oxidation reactions leading to production of NADPH occur in this early part of the pathway. Under certain metabolic conditions, the pentose phosphate pathway may end at this point, with the utilization of NADPH for reductive biosynthetic reactions and ribose 5-phosphate as a precursor for nucleotide synthesis. The overall equation may be written as

$$\text{Glucose 6-phosphate} + 2\text{NADP}^+ + \text{H}_2\text{O} \longrightarrow$$
$$\text{ribose 5-phosphate} + 2\text{NADPH} + 2\text{H}^+ + \text{CO}_2$$

Interconversions of Pentose Phosphate

In certain cells more NADPH is needed for reductive biosynthesis than ribose 5-phosphate for incorporation into nucleotides. A sugar rearrangement system (Figure 8.2) which leads to the formation of triose, tetrose, hexose, and heptose from the pentoses exists, thus creating a disposal mechanism for ribose 5-phosphate, as well as providing a reversible link between the pentose phosphate pathway and glycolysis, since certain of these sugars are common intermediates of both pathways. In order to allow the interconversions, another pentose phosphate, xylulose 5-phosphate, must first be formed through isomerization of ribulose 5-phosphate by the action of phosphopentose epimerase. As a consequence, these three pentose phosphates exist as an equilibrium mixture, and can then undergo further transformations catalyzed by transketolase and transaldolase. Both enzymes catalyze chain cleavage and transfer reactions involving the same group of substrates.

Transketolase, an enzyme requiring thiamine pyrophosphate (TPP) and Mg^{2+}, transfers a 2-carbon unit from xylulose 5-phosphate to ribose 5-phosphate, producing the 7-carbon sugar sedoheptulose and glyceraldehyde 3-phosphate, an intermediate of glycolysis. The essential feature of this reaction involves the transfer of a C_2 group, designated "active glycolaldehyde," from a suitable donor ketose to an acceptor aldose (another example will be encountered subsequently). A further transfer reaction, catalyzed by transaldolase, results in the recovery of the first hexose phosphate. In this reaction a 3-carbon unit (dihydroxyacetone) from sedoheptulose 7-phosphate is transferred to glyceraldehyde 3-phosphate, forming the tetrose, erythrose 4-phosphate, and fructose 6-phosphate, another intermediate of glycolysis. This reaction proceeds by a mechanism similar to that for fructose bisphosphate aldolase (Chapter 7), except that transaldolase is unable to react with or form free dihydroxyacetone or its phosphate.

In a third reaction, specific to this pathway, transketolase catalyzes the synthesis of fructose 6-phosphate and glyceraldehyde 3-phosphate from erythrose 4-phosphate and a second molecule of xylulose 5-phosphate. In this case, the 2-carbon unit is transferred from xylulose 5-phosphate to an acceptor 4-carbon sugar, now forming two glycolytic intermediates. The sum of these reactions is

2 Xylulose 5-phosphate + ribose 5-phosphate $\rightleftharpoons$
2 fructose 6-phosphate + glyceraldehyde 3-phosphate

Since xylulose 5-phosphate is derived from ribose 5-phosphate, the net reaction starting from ribose 5-phosphate is

3 Ribose 5-phosphate $\rightleftharpoons$
2 fructose 6-phosphate + glyceraldehyde 3-phosphate

diminished glucose 6-phosphate dehydrogenase activity. This enzymatic deficiency, which is usually undetected until administration of certain drugs, illustrates the interplay of heredity and environment on the production of disease. Enzyme absence is only one of several abnormalities that can affect enzyme activity, and others have been detected independent of drug administration. Increasingly, these variants can be distinguished from one another by certain clinical and biochemical differences.

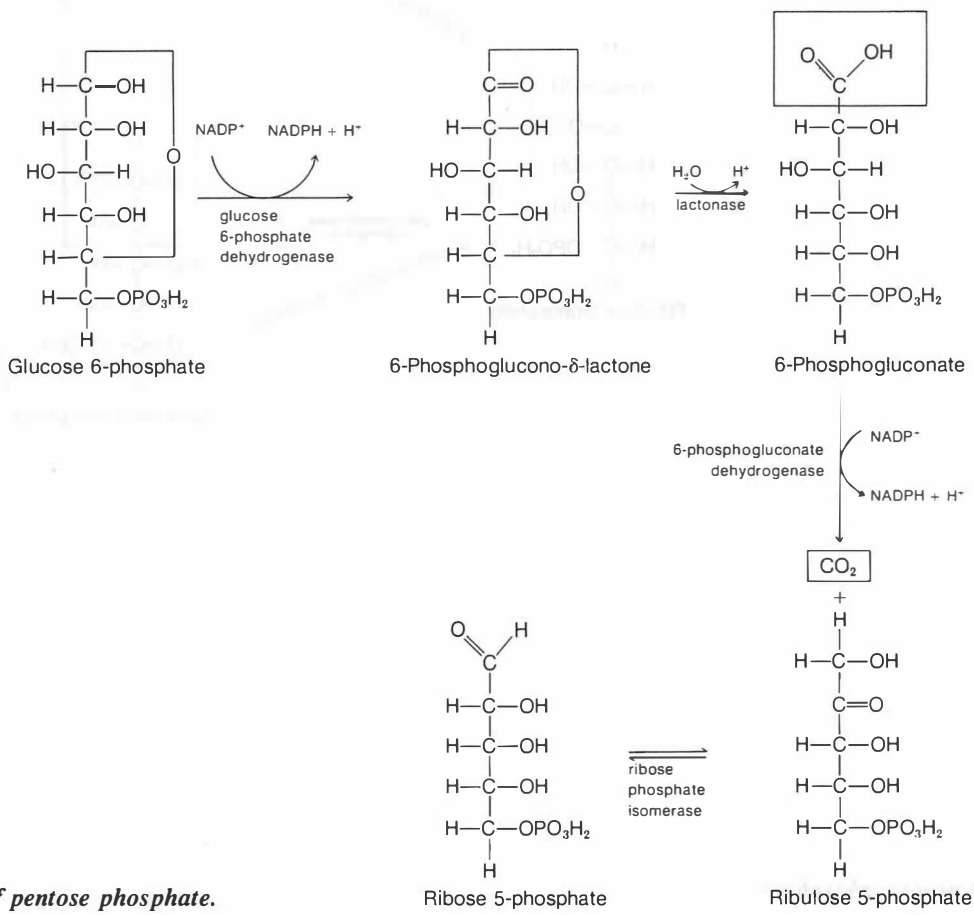


Figure 8.1
Formation of pentose phosphate.

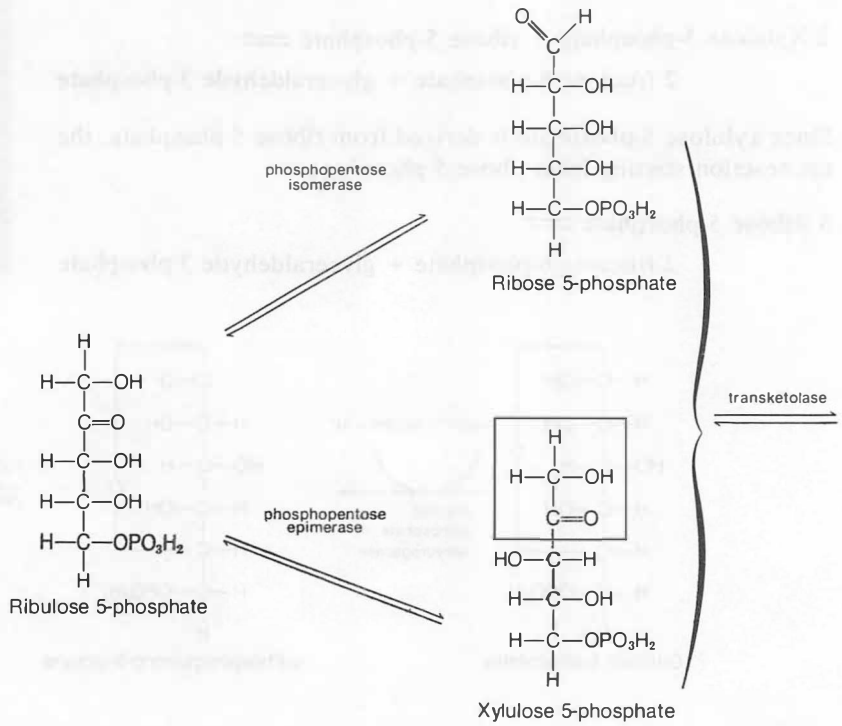
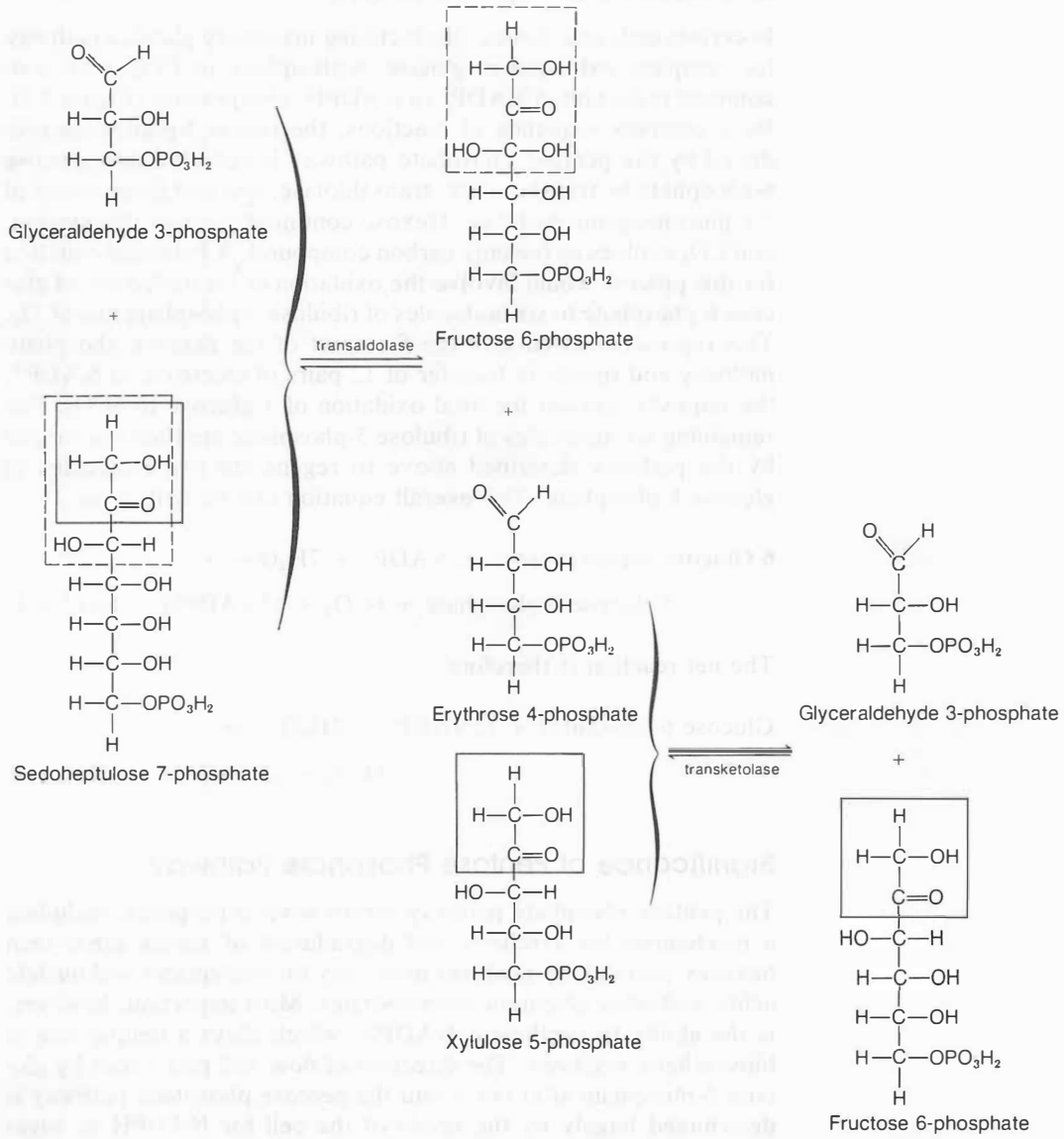


Figure 8.2
Interconversions of pentose phosphate.


Glyceraldehyde 3-phosphate
+
Sedoheptulose 7-phosphate
transaldolase
Fructose 6-phosphate
+
Erythrose 4-phosphate
Xylulose 5-phosphate
transketolase
Glyceraldehyde 3-phosphate
+
Fructose 6-phosphate

Therefore, excess ribose 5-phosphate, whether it arises from the initial oxidation of glucose 6-phosphate or from the degradative metabolism of nucleic acids, is effectively scavenged by conversion to intermediates that can enter the carbon flow of glycolysis.

Complete Oxidation of Glucose 6-phosphate to CO_2

In certain cells and tissues, like lactating mammary gland, a pathway for complete oxidation of glucose 6-phosphate to CO_2, with concomitant reduction of $NADP^+$ to NADPH, also prevails (Figure 8.3). By a complex sequence of reactions, the ribose 5-phosphate produced by the pentose phosphate pathway is recycled into glucose 6-phosphate by transketolase, transaldolase, and certain enzymes of the gluconeogenic pathway. Hexose continually enters this system, and CO_2 evolves as the only carbon compound. A balanced equation for this process would involve the oxidation of six molecules of glucose 6-phosphate to six molecules of ribulose 5-phosphate and $6CO_2$. This represents essentially the first part of the pentose phosphate pathway and results in transfer of 12 pairs of electrons to $NADP^+$, the requisite amount for total oxidation of 1 glucose to $6CO_2$. The remaining six molecules of ribulose 5-phosphate are then rearranged by the pathway described above to regenerate five molecules of glucose 6-phosphate. The overall equation can be written as

$$6 \text{ Glucose 6-phosphate} + 12NADP^+ + 7H_2O \longrightarrow 5 \text{ glucose 6-phosphate} + 6CO_2 + 12NADPH + 12H^+ + P_i$$

The net reaction is therefore

$$\text{Glucose 6-phosphate} + 12NADP^+ + 7H_2O \longrightarrow 6CO_2 + 12NADPH + 12H^+ + P_i$$

Significance of Pentose Phosphate Pathway

The pentose phosphate pathway serves several purposes, including a mechanism for synthesis and degradation of sugars other than hexoses, particularly pentoses necessary for nucleotides and nucleic acids, and other glycolytic intermediates. Most important, however, is the ability to synthesize NADPH, which plays a unique role in biosynthetic reactions. The direction of flow and path taken by glucose 6-phosphate after entry into the pentose phosphate pathway is determined largely by the needs of the cell for NADPH or sugar

intermediates. The situation in which more NADPH than ribose 5-phosphate is required has already been examined and results in a continuation of the pathway, leading to complete oxidation of glucose 6-phosphate to CO_2 and resynthesis of glucose 6-phosphate from ribose 5-phosphate. Alternatively, if more ribose 5-phosphate than NADPH is required, glucose 6-phosphate is converted to fructose 6-phosphate and glyceraldehyde 3-phosphate by the glycolytic pathway. Two molecules of fructose 6-phosphate and one molecule of glyceraldehyde 3-phosphate are then converted into three molecules of ribose 5-phosphate by a reversal of the transaldolase and transketolase reactions.

The distribution of the pentose phosphate pathway in the tissues of the body is consistent with its functions. As previously mentioned, it is present in erythrocytes for production of NADPH, which in turn is used to generate reduced glutathione, essential for maintenance of normal red cell structure. It is also active in tissues such as liver, mammary gland, testis, and adrenal cortex, which are active sites of fatty acid or steroid synthesis, processes that require

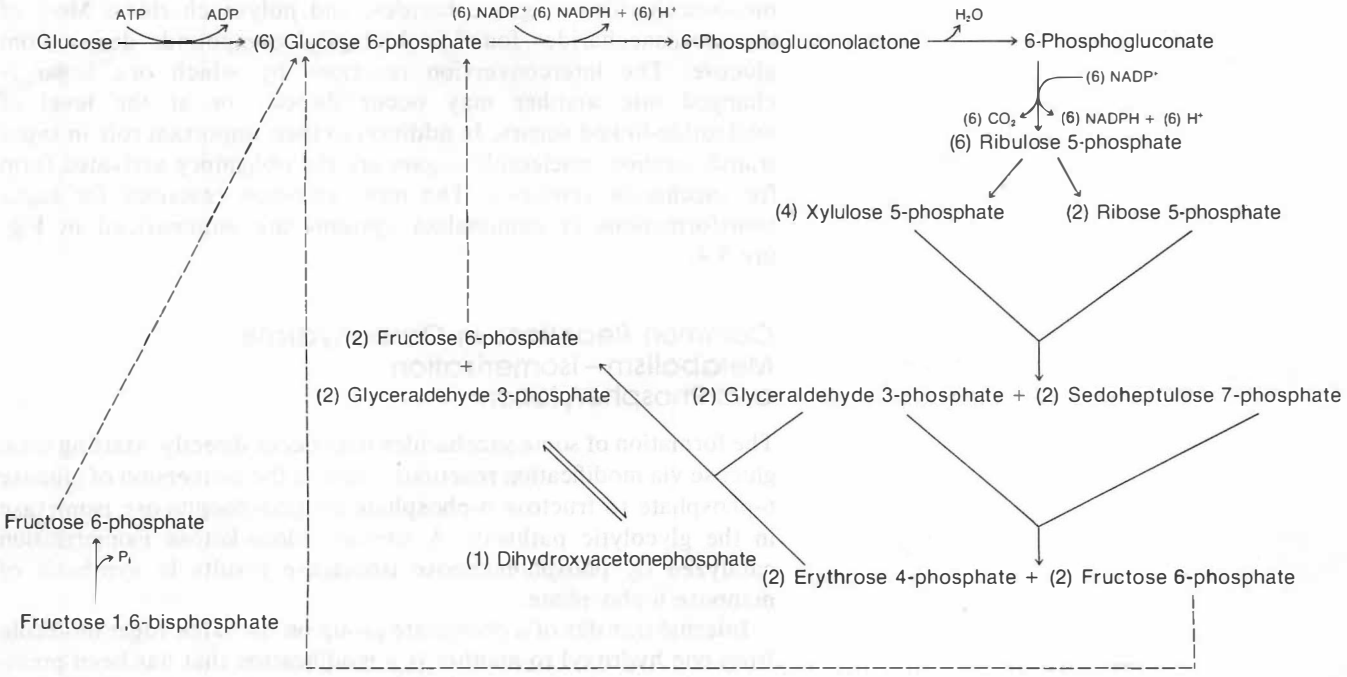


Figure 8.3
Pentose phosphate pathway.

the reducing power of NADPH. In contrast, in mammalian striated muscle, where little fatty acid or steroid synthesis occurs, there is no direct oxidation of glucose 6-phosphate through the pentose phosphate pathway. Rather, all catabolism proceeds via glycolysis and the TCA cycle. In some other tissues like liver, 20–30% of the CO_2 may arise from the pentose phosphate pathway, and the balance between glycolysis and the pentose phosphate pathway depends on the metabolic requirements of the cell.

8.2 SUGAR INTERCONVERSIONS AND NUCLEOTIDE SUGAR FORMATION

In preceding discussions, the general principles of carbohydrate metabolism, specifically those involving glucose, were considered. Now we shall examine certain aspects of the metabolism of other monosaccharides, oligosaccharides, and polysaccharides. Most of the monosaccharides found in biological compounds derive from glucose. The interconversion reactions by which one sugar is changed into another may occur directly or at the level of nucleotide-linked sugars. In addition to their important role in sugar transformation, nucleotide sugars are the obligatory activated form for saccharide synthesis. The most common reactions for sugar transformations in mammalian systems are summarized in Figure 8.4.

Common Reactions in Carbohydrate Metabolism—Isomerization and Phosphorylation

The formation of some saccharides may occur directly, starting from glucose via modification reactions, such as the conversion of glucose 6-phosphate to fructose 6-phosphate by phosphoglucose isomerase in the glycolytic pathway. A similar aldose-ketose isomerization catalyzed by phosphomannose isomerase results in synthesis of mannose 6-phosphate.

Internal transfer of a phosphate group on the same sugar molecule from one hydroxyl to another is a modification that has been previously described. Glucose 1-phosphate, resulting from enzymatic hydrolysis of glycogen, is converted to glucose 6-phosphate for entry into the glycolytic pathway by phosphoglucomutase. Galactose may

also be phosphorylated directly to galactose 1-phosphate by a galactokinase and mannose to mannose 6-phosphate by a mannokinase; the latter may equilibrate with fructose 6-phosphate as discussed above. Similarly, free fructose, an important dietary constituent, may be phosphorylated in the liver to fructose 1-phosphate by a special fructokinase. However, no mutase exists to interconvert fructose 1-phosphate and fructose 6-phosphate, nor can phosphofructokinase synthesize fructose diphosphate from fructose 1-phosphate. Rather, a specific adolase cleaves fructose 1-phosphate to dihydroxyacetone phosphate (DHAP), which enters the glycolytic pathway directly, and glyceraldehyde which must first be reduced to glycerol, phosphorylated, and then reoxidized to DHAP. Lack of this aldolase may lead to fructose intolerance (see Clin. Corr. 8.2).

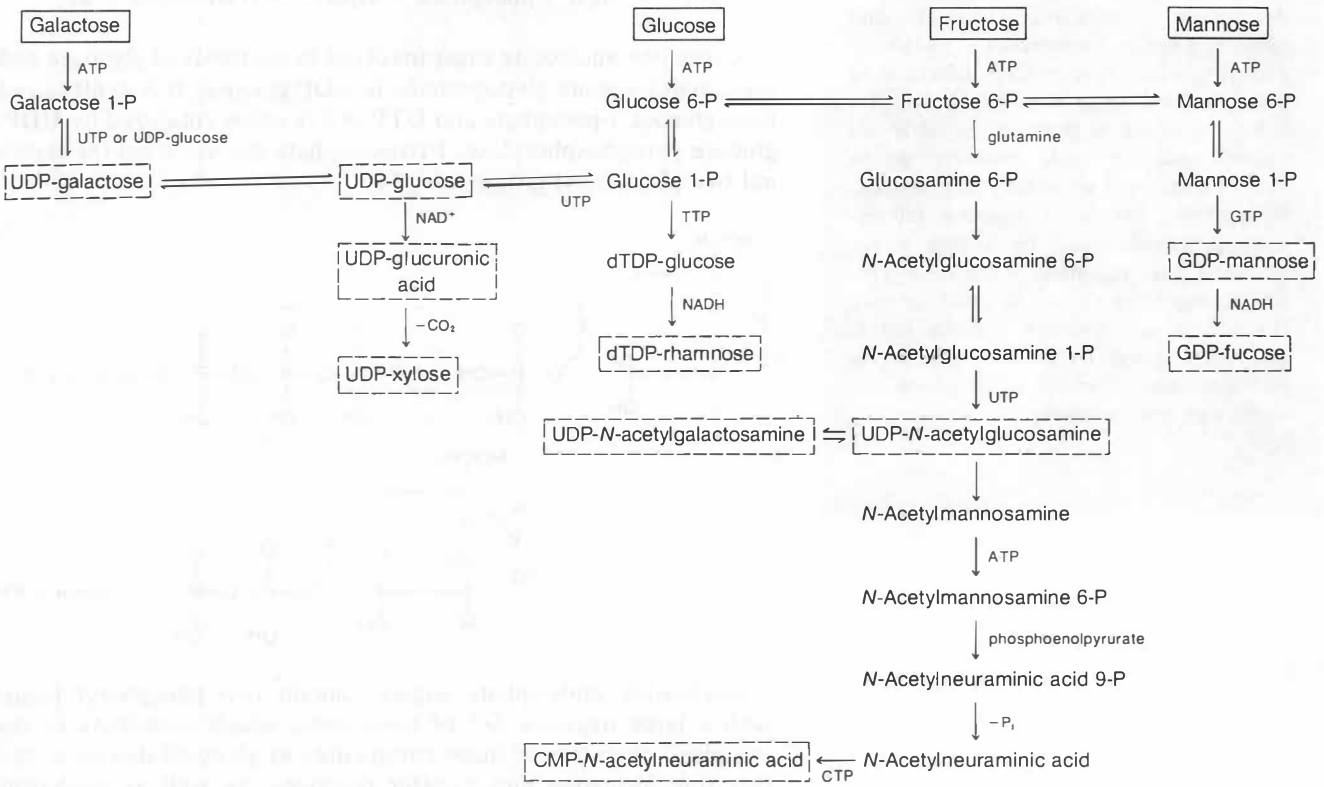


Figure 8.4
Pathways of formation of sugar nucleotides and interconversions of some hexoses.

CLIN. CORR. 8.2 ESSENTIAL FRUCTOSURIA AND FRUCTOSE INTOLERANCE: DEFICIENCY OF FRUCTOKINASE AND FRUCTOSE 1-PHOSPHATE ALDOLASE, RESPECTIVELY

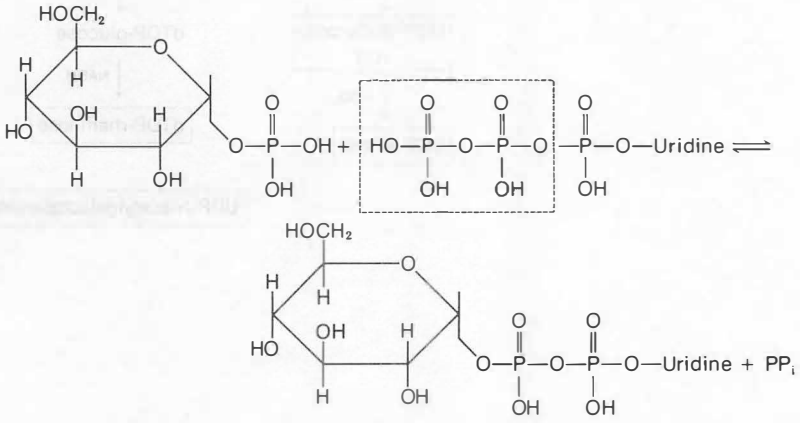
Fructose may account for 30 to 60% of the total carbohydrate intake of mammals. It is metabolized more than 50% by a specific fructose pathway. The first enzyme in this pathway, fructokinase, is deficient in essential fructosuria. This disorder is a benign asymptomatic metabolic anomaly, which appears to be inherited as an autosomal recessive. Biochemically, following intake of fructose, blood levels appear unusually high; however, 80 to 90% of fructose is eventually metabolized. In contrast, hereditary fructose intolerance is characterized by severe hypoglycemia after ingestion of fructose. Prolonged ingestion in young children may lead to a chronic condition or death. In this disorder fructose 1-phosphate aldolase is deficient, and fructose 1-phosphate accumulates intracellularly.

Formation of Nucleotide Sugars

Most other sugar transformation reactions require the prior conversion into *nucleotide*-linked sugars. Nucleotides are phosphoric acid esters of nucleosides; nucleosides are composed of a pentosyl derivative of a purine or pyrimidine base. Formation of *nucleoside diphosphate* sugar involves the reaction of hexose 1-phosphate and *nucleoside* triphosphate, catalyzed by a pyrophosphorylase. These reactions are readily reversible. However, in vivo pyrophosphate is rapidly hydrolyzed irreversibly by an inorganic pyrophosphatase, thereby driving the synthesis of nucleotide sugars. These reactions are summarized as follows:

$$\text{NTP} + \text{sugar 1-phosphate} \rightleftharpoons \text{NDP-sugar} + PP_i$$
$$PP_i + H_2O \longrightarrow 2P_i$$
$$\overline{\text{NTP} + \text{sugar 1-phosphate} + H_2O \longrightarrow \text{NDP-sugar} + 2P_i}$$

A common nucleotide sugar involved in synthesis of glycogen and presumably certain glycoproteins is UDP-glucose. It is synthesized from glucose 1-phosphate and UTP in a reaction catalyzed by UDP-glucose pyrophosphorylase. Pyrophosphate derives from the terminal two phosphoryl groups of UTP.



Nucleoside diphosphate sugars contain two phosphoryl bonds with a large negative ΔG of hydrolysis, which contribute to the energized character of these compounds as glycosyl donors in further transformation and transfer reactions, as well as conferring specificity on the enzymes catalyzing these reactions. This is discussed in greater detail in Section 8.3. UDP usually serves as the glycosyl carrier in higher animals; however, ADP, GDP, and CMP have also been shown to act as carriers in other reactions. As previ-

ously mentioned, many of the sugar transformation reactions occur only at the level of nucleotide sugars. Some of these modification reactions involving nucleotide sugars include epimerization, oxidation, decarboxylation, reduction, and rearrangement.

Epimerization

One of the most common types of these reactions in carbohydrate metabolism is epimerization. For example, the reversible conversion of glucose to galactose in animals occurs by enzymatic epimerization of UDP-glucose to UDP-galactose, catalyzed by UDP-glucose epimerase. The UDP-galactose that participates in the epimerization reaction is an important intermediate in the metabolism of free galactose, derived from the hydrolysis of lactose in the intestinal tract, and may be formed in either of two ways. Galactose can first be phosphorylated by galactokinase and ATP to yield galactose 1-phosphate:

$$\text{Galactose} + \text{ATP} \rightleftharpoons \text{galactose 1-phosphate} + \text{ADP}$$

An enzyme, galactose 1-phosphate uridylyltransferase, transforms galactose 1-phosphate into UDP-galactose by displacing glucose 1-phosphate from UDP-glucose:

$$\text{UDP-glucose} + \text{galactose 1-phosphate} \rightleftharpoons \text{UDP-galactose} + \text{glucose 1-phosphate}$$

A hereditary disorder, galactosemia, results from the absence of this uridylyltransferase (Clin. Corr. 8.3). Alternatively, UDP-galactose can be formed directly from galactose 1-phosphate and UTP, catalyzed by UDP-galactose pyrophosphorylase:

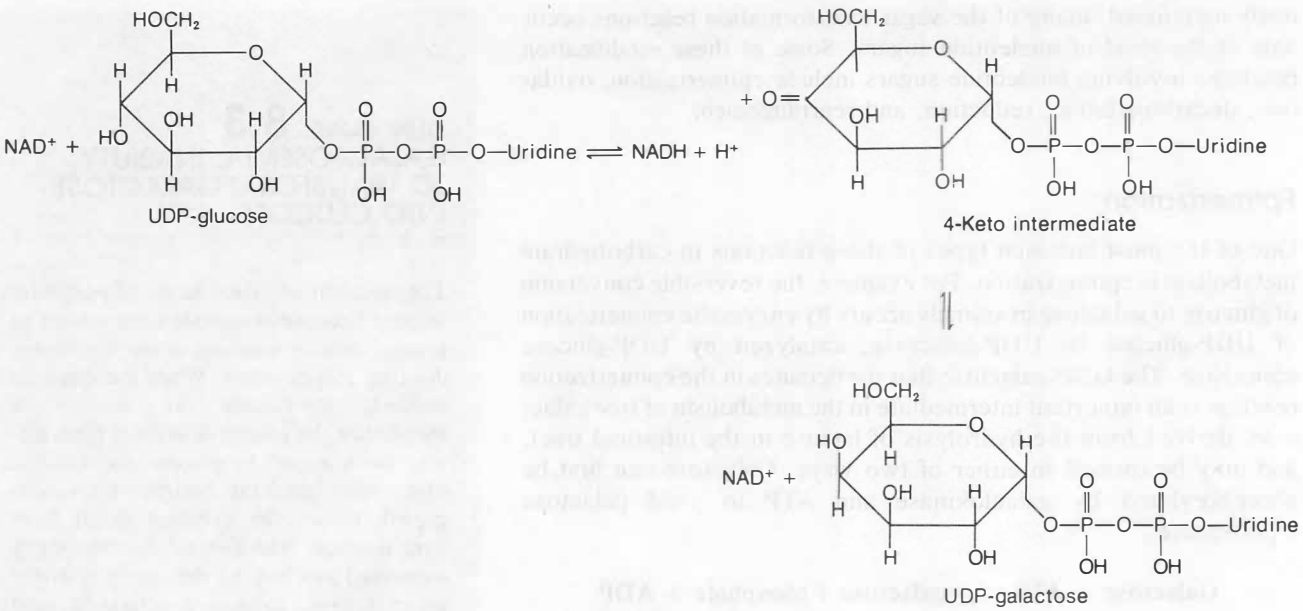
$$\text{Galactose 1-phosphate} + \text{UTP} \rightleftharpoons \text{UDP-galactose} + \text{PP}_i$$

However, this alternative pathway for galactose 1-phosphate metabolism only develops in later life.

An NAD^+-dependent galactose 4-epimerase converts an equatorial hydroxyl at C-4 in glucose to an axial one in galactose in the epimerization of these two sugars. Most likely, NAD^+ accepts the hydrogen atom at C-4, and a 4-keto intermediate is formed. Inversion of the hydroxyl group then occurs when NADH transfers its hydrogen to the other side of C-4. A combination of these reactions allows an efficient transformation of galactose derived from the diet into glucose 1-phosphate, which can then be further metabolized by

CLIN. CORR. 8.3 GALACTOSEMIA: INABILITY TO TRANSFORM GALACTOSE INTO GLUCOSE

The reactions of galactose are of particular interest because in man they are subject to genetic defects resulting in the hereditary disorder galactosemia. When the defect is present, individuals are unable to metabolize the galactose derived from lactose (milk sugar) to glucose metabolites, often with resultant cataract formation, growth failure, or eventual death from liver damage. The genetic disturbance is expressed as a cellular deficiency of either galactokinase, causing a relatively mild disorder, or galactose 1-phosphate uridylyltransferase, resulting in a more severe form. Since the alternate enzyme for forming UDP-galactose directly from galactose 1-phosphate does not become active until later years in life, infants with galactosemia are unable to metabolize galactose by either pathway. The development of this second enzyme does account for the fact that galactosemic patients tend to become resistant to galactose as they mature. Thus, if maintained on a galactose-free diet during infancy, serious damage may be avoided.



previously described pathways. Alternatively, the 4-epimerase can operate in the reverse direction when UDP-galactose is needed for biosynthesis. Other 4-epimerases are known such as those that convert UDP-*N*-acetylglucosamine to UDP-*N*-acetylgalactosamine, or UDP-xylose to UDP-arabinose. Presumably these operate by a mechanism similar to the UDP-galactose 4-epimerase.

More recent work has shown that epimerization reactions are not exclusively restricted to nucleotide-linked sugars but may also occur at the polymer level. Thus D-mannuronic acid is epimerized to L-guluronic acid after incorporation into alginic acid (a polyglycosyluronate compound produced by seaweed and certain bacteria), and D-glucuronic acid is epimerized to L-iduronic acid after incorporation into heparin and dermatan sulfate, which is discussed later.

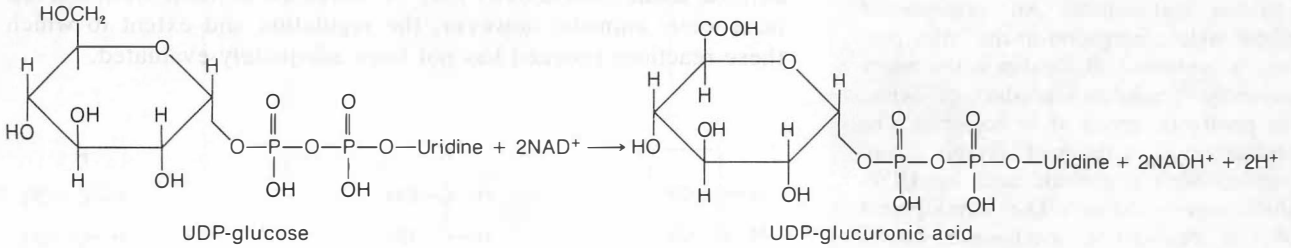
Glucuronic Acid Metabolism

Oxidation and reduction interconversions also result in formation of many additional sugars. One of the most important is formation of glucuronic acid, which serves as a precursor of L-ascorbic acid in those animals which synthesize vitamin C, is converted to L-xylulose, the ketopentose excreted by humans with essential pentosuria (Clin. Corr. 8.4), and participates in physiological processes

CLIN. CORR. 8.4 PENTOSURIA: DEFICIENCY OF XYLITOL DEHYDROGENASE

The glucuronic acid oxidation pathway presumably is not essential to human carbohydrate metabolism, since individuals in whom the pathway is blocked suffer no ill effects. This metabolic variation, called idiopathic pentosuria, results from reduced activity of NADP-linked xylitol dehydrogenase, the enzyme that catalyzes the reduction of xylulose to xylitol. Hence affected individuals excrete large amounts of pentose into the urine especially following intake of glucuronic acid.

of urinary detoxification by production of glucuronide conjugates (Clin. Corr. 8.5). Glucuronic acid is formed by oxidation of UDP-glucose catalyzed by UDP-glucose dehydrogenase. The UDP-



glucose dehydrogenase-catalyzed reaction is important in the overall pathway for conversion of glucose to glucuronic acid, and most likely follows the scheme outlined in Figure 8.5. UDP-glucuronic acid may then be epimerized to UDP-galacturonic acid. In a similar manner, GDP-mannose is oxidized to GDP-mannuronic acid, which may undergo epimerization to GDP-guluronic acid.

Following the formation of free glucuronic acid, it is further metabolized by reduction with NADPH to L-gulonic acid (Figure 8.6). Gulonic acid is converted by a two-step process through L-gulonolactone to L-ascorbic acid (vitamin C) in plants and most higher animals. Man, other primates, and the guinea pig lack the enzyme that converts L-gulonolactone to L-ascorbic acid and therefore must satisfy their needs for ascorbic acid by ingestion. Gulonic acid may also be oxidized through 3-ketogulonic acid and decarboxylated to L-xylulose. L-Xylulose is in turn converted by reduction to xylitol, reoxidized to D-xylulose and phosphorylated with ATP and an appropriate kinase to xylulose 5-phosphate. The latter compound may then re-enter the pentose phosphate pathway described previously. This complex catabolic pathway for glucuronic acid represents another shunt pathway for oxidation of glucose. It should be

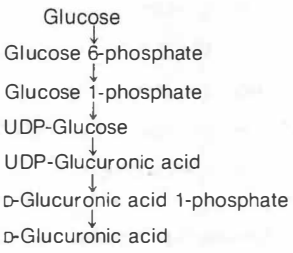


Figure 8.5
Biosynthesis of D-glucuronic acid.

CLIN. CORR. 8.5 GLUCURONIC ACID: PHYSIOLOGICAL SIGNIFICANCE OF GLUCURONIDE FORMATION

The biological significance of glucuronic acid extends to its ability to conjugate with certain endogenous and exogenous substances, forming a group of compounds collectively termed glucuronides in a reaction catalyzed by UDP-glucuronyltransferase. Conjugation of a compound with glucuronic acid produces a strongly acidic compound that is more water-soluble at physiological pH than the precursor and therefore may alter its metabolism, transport, or excretion properties. Glucuronide formation is important in

several processes, including drug detoxification, steroid glucurono conjugation, and bilirubin metabolism. An example of glucuronide conjugation in the latter process is presented. Bilibrubin is the major metabolic breakdown product of heme, the prosthetic group of hemoglobin. The central step in excretion of bilirubin is conjugation with glucuronic acid by UDP-glucuronyltransferase. The development of this conjugating mechanism occurs gradually and may take several days to 2 weeks after birth to be active in humans. Although not from a single cause, "physiological jaundice of the newborn" may result from the inability of the neonatal liver to form bilirubin glucuronide at a rate comparable to that of bilirubin production. A defect in glucuronide synthesis has been found in a mutant strain of Wistar ("Gunn") rats, due to a deficiency of UDP-glucuronyltransferase and results in hereditary hyperbilirubinemia. In human beings a similar defect is found in congenital familial nonhemolytic jaundice (Crigler–Najjar syndrome). Patients with this condition are also unable to conjugate foreign compounds efficiently with glucuronic acid.

noted that, in contrast to other pathways of carbohydrate metabolism in which only phosphate esters participate, these reactions also involve free sugars. Evidence suggests that this pathway operates in adipose tissue, and activity may be increased in tissue from starved or diabetic animals; however, the regulation and extent to which these reactions proceed has not been adequately evaluated.

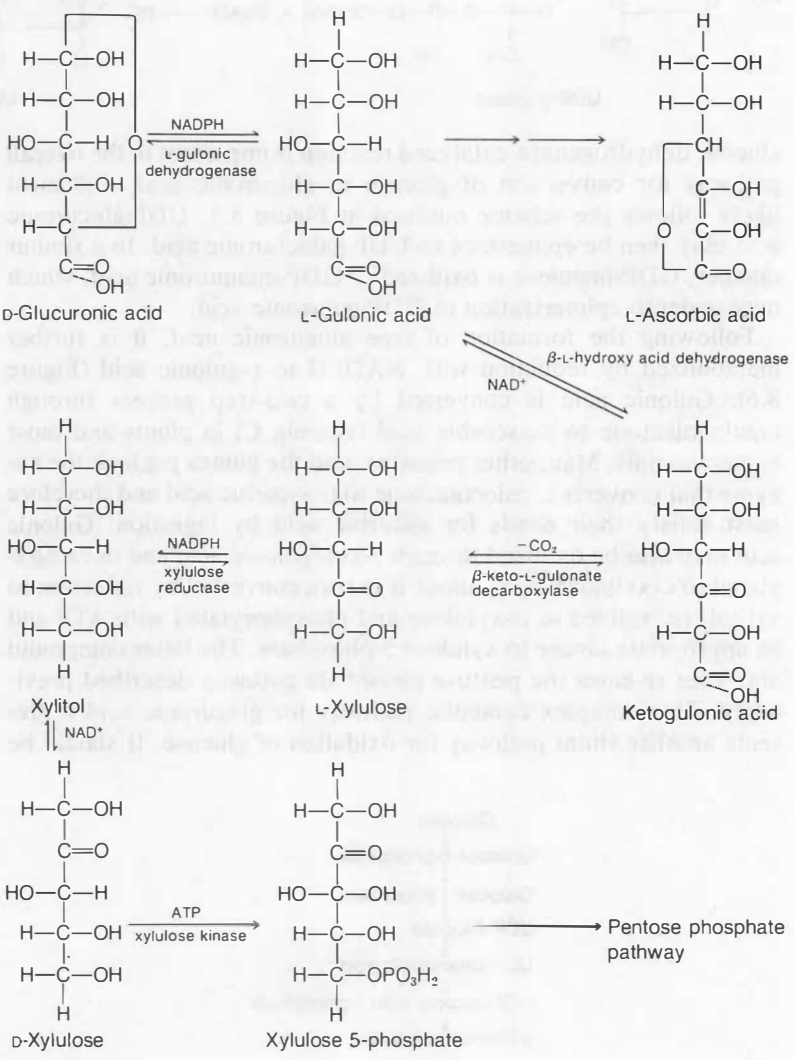


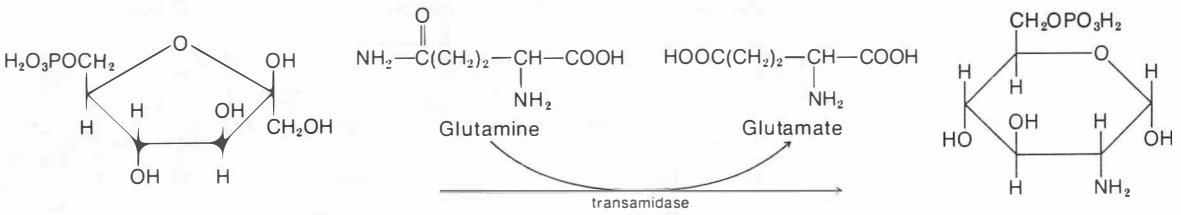
Figure 8.6
Glucuronic acid oxidation pathway.

Decarboxylation, Oxidoreduction, and Transamination

Decarboxylation, which is an important mechanism for degrading sugars one carbon atom at a time, has also previously been encountered in the major metabolic pathways. The only known decarboxylation of a nucleotide sugar is the conversion of UDP-glucuronic acid to UDP-xylose, necessary for synthesis of proteoglycans (Section 8.6). UDP-xylose is a potent inhibitor of UDP-glucose dehydrogenase, which oxidizes UDP-glucose to UDP-glucuronic acid (Figure 8.4). Thus the level of these nucleotide sugar precursors is regulated by this sensitive feedback mechanism.

Deoxyhexoses and dideoxyhexoses are also synthesized, while the sugars are attached to nucleoside diphosphates, by a multistep process. For example, L-rhamnose is synthesized from glucose by a series of oxidation–reduction reactions starting with dTDP-glucose and yielding dTDP-rhamnose, catalyzed by oxidoreductases. Presumably, similar reactions account for the synthesis of GDP-fucose from GDP-mannose, as well as for various dideoxyhexoses.

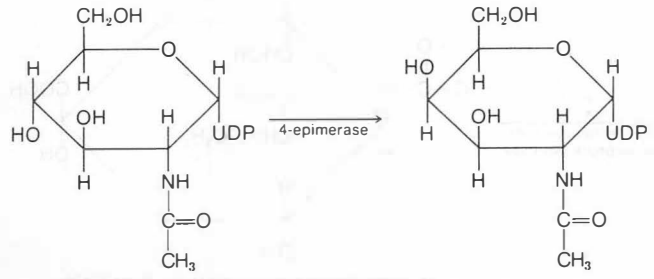
Formation of amino sugars, often found as constituents of antibiotics, occurs by transamidation. For example, synthesis of glucosamine 6-phosphate occurs by the reaction of fructose 6-phosphate with glutamine. Glucosamine 6-phosphate may then be



Fructose 6-phosphate

Glucosamine 6-phosphate

N-acetylated, forming *N*-acetylglucosamine 6-phosphate, followed by isomerization to *N*-acetylglucosamine 1-phosphate. This latter sugar is converted to UDP-*N*-acetylglucosamine by reactions similar to those of UDP-glucose synthesis. UDP-*N*-acetylglucosamine, a



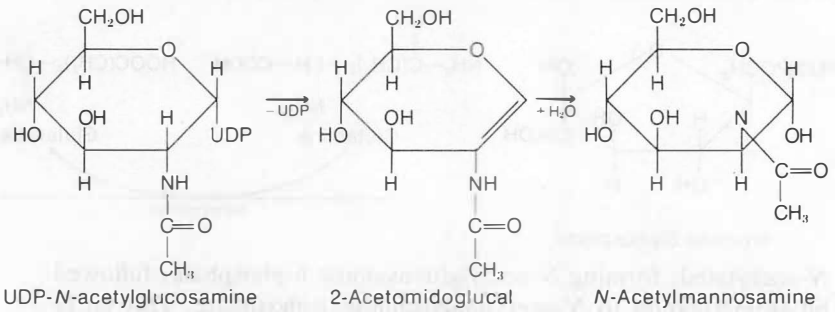
UDP-*N*-acetylglucosamine

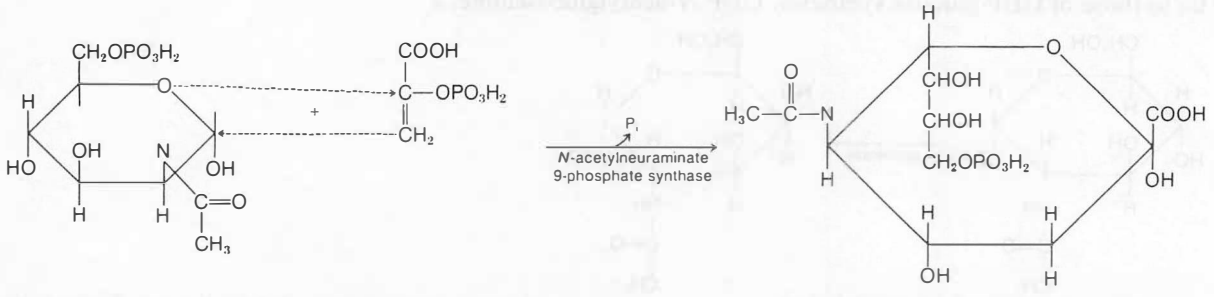
UDP-*N*-acetylgalactosamine

precursor of glycoprotein synthesis, may be epimerized to UDP-*N*-acetylgalactosamine, necessary for proteoglycan synthesis. The first reaction in hexosamine synthesis, the fructose 6-phosphate: glutamine transamidase reaction, is under negative feedback control by UDP-*N*-acetylglucosamine; thus, the synthesis of both nucleotide sugars is regulated (Figure 8.4). This regulation is meaningful in certain tissues such as skin, in which this pathway may involve up to 20% of the glucose flux.

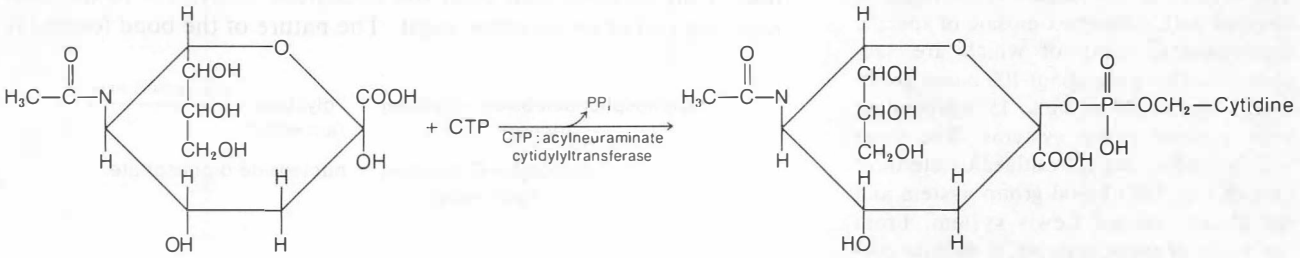
Synthesis of Sialic Acid

Another product derived from UDP-*N*-acetylglucosamine is CMP-*N*-acetylneuraminic acid, one of a family of 9-carbon sugars, called sialic acids. The first reaction in this complex pathway involves epimerization of UDP-*N*-acetylglucosamine by a 2-epimerase to *N*-acetylmannosamine, concomitant with elimination of UDP. Since the monosaccharide product is no longer bound to nucleotide, this epimerization is clearly different from those previously encountered. Most likely, this 2-epimerase reaction proceeds by a *trans* elimination of UDP, with formation of the unsaturated intermediate, 2-acetomidoglucal. In mammalian tissues *N*-acetyl-





mannosamine is phosphorylated with ATP to *N*-acetylmannosamine 6-phosphate, which then condenses with phospho*enol*pyruvate to form *N*-acetylneuraminic acid 9-phosphate. This product is cleaved by a phosphatase and activated by CTP to form the CMP derivative, CMP-*N*-acetylneuraminic acid. This is an unusual nucleotide sugar containing only one phosphate group, and is formed by a reaction that is irreversible. *N*-Acetylneuraminic acid



N-Acetylneuraminic acid

CMP-*N*-acetylneuraminic acid

is a precursor of other sialic acid derivatives, some of which evolve by modification of *N*-acetyl to *N*-glycolyl or *O*-acetyl after incorporation into glycoprotein.

8.3 COMPLEX CARBOHYDRATES

In addition to their role as the main source of energy in living organisms, monosaccharides are often found as components of more complex macromolecules like oligo- and polysaccharides, glycoproteins, glycolipids, and proteoglycans. Certain oligosaccharides, covalently linked to protein or lipids, form the major structural components of bacterial cell walls, and include peptidoglycan, teichoic acids, and lipopolysaccharides. Chitin, a linear homopolymer of *N*-acetylglucosamine residues, is the predominant organic structural component of the exoshells of the invertebrates, as well as most fungi, many algae, and some yeast. Cellulose, an unbranched polymer of glucose residues, is the major structural component in the plant kingdom. In higher animals some of the complex carbohydrate molecules are also predominantly structural elements found in ground substance filling the extracellular space in tissues and as components of cell membranes. Increasingly though, more dynamic functions for these complex macromolecules, such as recognition markers and as determinants of biological specificity, are being discovered. The following discussion is limited to the chemistry and

CLIN. CORR. 8.6 BLOOD GROUP SUBSTANCES

The surface of the human erythrocyte is covered with a complex mosaic of specific determinants, many of which are saccharides. There are about 100 blood group determinants, belonging to 15 independent human blood group systems. The most widely studied are the antigenic determinants of the ABO blood group system and the closely related Lewis system. From the study of these systems, a definite correlation was established between gene activity as it relates to specific glycosyltransferase synthesis and oligosaccharide structure. The genetic variation is achieved through specific glycosyltransferases responsible for synthesis of the heterosaccharide determinants. For example, the *H* gene codes for a fucosyltransferase, which adds fucose to a peripheral galactose in the heterosaccharide precursor. A and B blood group specificities arise by addition of either *N*-acetylgalactosamine or galactose to the nonreducing terminal galactose of the *H*-specific oligosaccharide. The Lewis (*Le*) gene codes for another fucosyltransferase, which adds fucose to a peripheral *N*-acetylglucosamine residue in the precursor. Absence of the *H* gene gives rise to the Le^a specific determinant, whereas in the absence of both the *H* and *Le* genes, the interaction product responsible for the Le^b specificity is found. The elucidation of the structures of these oligosaccharide determinants represents a milestone in carbohydrate chemistry. This knowledge has proved useful not only for medicine and biology, but also for legal and historical purposes. For example, tissue dust containing complex carbohydrates

biology of those complex carbohydrates found in animal tissues and fluids.

In complex carbohydrate-containing molecules, sugars are linked to other sugars by glycosidic bonds, which are formed by specific glycosyltransferases. Energy is required for synthesis of a glycosidic bond and is made available through the use of nucleotide sugars as donor substrates. A glycosyltransferase reaction proceeds by donation of the glycosyl unit from the nucleotide derivative to the nonreducing end of an acceptor sugar. The nature of the bond formed is

$$\underset{\text{(donor)}}{\text{Diphosphonucleoside—glycose}_1} + \underset{\text{(acceptor)}}{\text{glycose}_2} \xrightarrow{\text{glycosyltransferase}} \underset{\text{(glycoside)}}{\text{glycosyl}_1\text{—}O\text{-glycose}_2} + \text{nucleoside diphosphate}$$

determined by the specificity of an individual glycosyltransferase, which is unique for the sugar acceptor, the sugar transferred, and the linkage formed. Thus polysaccharide synthesis is controlled by a nontemplate mechanism (see Chapter 17) in which genes code for specific glycosyltransferases.

At least 40 different glycosidic bonds have been identified in mammalian oligosaccharides and about 15 additional ones in connective tissue polysaccharides. The number of possible linkages is even greater and arises both from the diversity of monosaccharides covalently bonded and from the formation of both α and β linkages, with each of the available hydroxyl groups on the acceptor saccharide. The large and diverse number of molecules that can be generated suggests that oligosaccharides have the potential for great informational content. In fact, it is known that the specificity of many biological molecules is determined by the nature of the composite sugar residues. For example, the specificity of the major blood types is determined by sugars (see Clin. Corr. 8.6). *N*-Acetylgalactosamine is the immunodeterminant of blood type A and galactose of blood type B. Removal of *N*-acetylgalactosamine from type A erythrocytes, or of galactose from type B erythrocytes, will convert both to type O erythrocytes. Increasingly, other examples of sugars as determinants of specificity for cell surface receptor and lectin interactions, targeting of cells to certain tissues, and survival or clearance from the circulation of certain molecules, are being recognized.

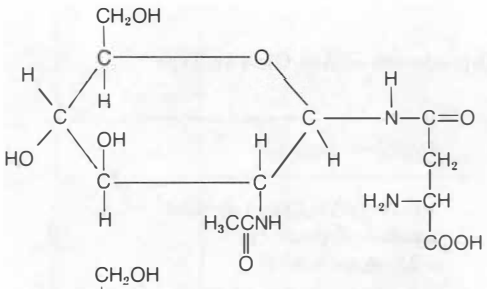
All the glycosidic bonds that have been identified in biological compounds are degraded by specific hydrolytic enzymes, glycosidases. In addition to being valuable tools for the structural elucidation of oligosaccharides, recent interest in this class of enzymes stems from the fact that many genetic diseases of complex

carbohydrate metabolism result from defects in glycosidases (see Clin. Corrs. 8.7 and 8.8 and Tables 8.1 and 8.2).

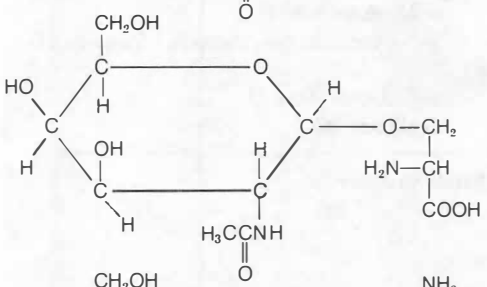
has been used in serological analysis to establish the blood group of Tutankhamen and his probable ancestral background.

8.4 GLYCOPROTEINS

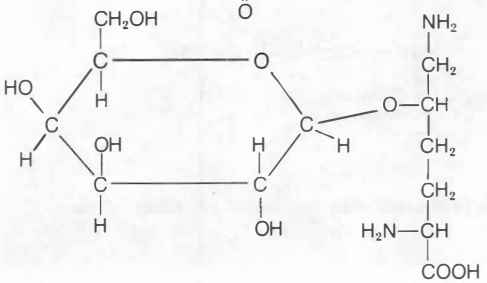
Glycoproteins have been defined restrictively as conjugated proteins containing as a prosthetic group one or more saccharides lacking a serial repeat unit and bound covalently to a peptide chain. This definition excludes proteoglycans, which are discussed in later sections. At present, the structures of relatively few oligosaccharide components have been completely elucidated. The microheterogeneity of glycoproteins, arising from incomplete synthesis or partial degradation, makes structural analyses extremely difficult. However, certain generalities about the structure of glycoproteins have emerged.



N-Glycosyl linkage to asparagine



O-Glycosyl linkage to serine or threonine



O-Glycosyl linkage to hydroxylysine

Figure 8.7
Structure of three major glycopeptide bonds.

Structure of Glycoprotein

The covalent linkage of sugars to the peptide chain is a central part of glycoprotein structure, and only a limited number of bonds are found. The three major types of glycopeptide bonds, as shown in Figure 8.7, are *N*-glycosyl to asparagine (Asn), *O*-glycosyl to serine (Ser), or threonine (Thr) and *O*-glycosyl to hydroxylysine. The latter linkage, representing the carbohydrate side chains of either a single galactose or the disaccharide glucosylgalactose covalently bonded to hydroxylysine, is generally confined to the collagens. The other two linkages occur in a wide variety of glycoproteins. Of the three major types, only the *O*-glycosidic linkage to serine or threonine is labile to alkali cleavage. By this procedure two types of oligosaccharides (simple and complex) are released. Examination of the simple class from porcine submaxillary mucins reveals some general structural features. A core structure exists, consisting of galactose (Gal) linked β-(1 → 3) to *N*-acetylgalactosamine (GalNAc) *O*-glycosidically

CLIN. CORR. 8.7 ASPARTYLGLYCOSYLAMINURIA: ABSENCE OF 4-L-ASPARTYLGLYCOSAMINE AMIDOHYDROLASE

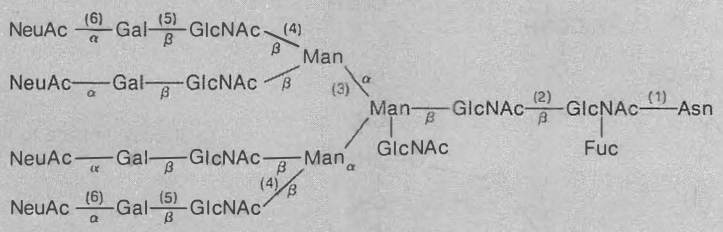
A group of human inborn errors of metabolism involving storage of glycolipids, glycopeptides, mucopolysaccharides, and oligosaccharides exists. These diseases are caused by defects in glycosidase activity, which prevents the catabolism of oligosaccharides. The disorders involve accumulation in tissues and urine of structures derived from incomplete degradation of the oligosaccharides, and may be accompanied by skeletal abnormalities, renal, hepatic, or cardiovascular defects, or severe mental retardation. One disorder resulting from a defect in catabolism of asparagine-*N*-acetylglucosamine-linked oligosaccharides is aspartylglycosylaminuria. A deficiency in the enzyme 4-L-aspartylglycosylamine amidohydrolase allows the accumulation of aspartylglucosamine-linked structures (Table 8.1).

Other disorders have been described involving accumulation of oligosaccharides derived from both glycoproteins and glycolipids, which may share common oligosaccharide structures (Table 8.1). Examples of genetic diseases include mannosidosis (L-mannosidase), G_{M2} gangliosidosis variant O (Sandhoff-Jatzkewitz disease; β-*N*-acetylhexosaminidases A and B), and G_{M1} gangliosidosis (β-galactosidase). (See also Chapter 10.)

Table 8.1 Enzymic Defects in Degradation of Asn-GlcNAc Type Glycoproteins[a]

Disease	*Deficient Enzyme*[b]
Aspartylglycosylaminuria	4-L-Aspartylglycosylamine amidohydrolase (1)
Mannosidosis	α-Mannosidase (3)
G_{M2}-Gangliosidosis variant O (Sandhoff–Jatzkewitz disease)	β-*N*-Acetylhexosamidases (A and B) (4)
G_{M1}-Gangliosidosis	β-Galactosidase (5)
Mucolipidosis I and II	Sialidase (6)

[a] A typical Asn-GlcNAc oligosaccharide structure.



[b] The numbers in parentheses refer to the enzymes that hydrolyze those bonds.

linked to serine or threonine residues. Residues of L-fucose (Fuc), sialic acid (NeuAc), and another *N*-acetylgalactosamine are found at the nonreducing periphery of this class of glycopeptides. The general structure of this type of glycopeptide is as follows:

$$\begin{array}{l} \text{GalNAc} \xrightarrow{1,3} \text{Gal} \xrightarrow{1,3} \text{GalNAc} \rightarrow O\text{-Ser/Thr} \\ \quad\quad\quad\quad\ \uparrow 1,2 \quad\quad\ \uparrow 2,6 \\ \quad\quad\quad\quad\ \text{Fuc} \quad\quad\ \text{NeuAc} \end{array}$$

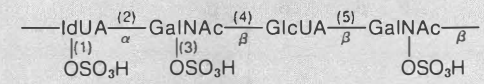
More complex heterosaccharides are also linked to peptide via serine or threonine and are exemplified by the blood group substances. The study of these determinants has shown how complex and variable these structures are, as well as how the oligosaccharides of cell surfaces are assembled and how that assembly pattern is genetically determined. An example of how oligosaccharide structures on the surface of red blood cells determine blood group specificity is presented in Clin. Corr. 8.6.

Table 8.2 Enzymic Defects in the Mucopolysaccharidoses

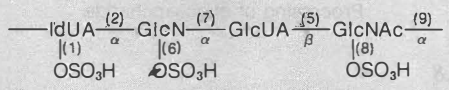
Disease	*Accumulated Products*[a]	*Deficient Enzyme*[b]
Hunter	Heparan sulfate Dermatan sulfate	Iduronate sulfatase (1)
Hurler + Scheie	Heparan sulfate Dermatan sulfate	α-L-Iduronidase (2)
Maroteaux–Lamy	Dermatan sulfate	*N*-Acetylgalactosamine (3) sulfatase
Mucolipidosis VII	Heparan sulfate Dermatan sulfate	β-Glucuronidase (5)
Sanfilippo A	Heparan sulfate	Heparan sulfamidase (6)
Sanfilippo B	Heparan sulfate	*N*-Acetylglucosaminidase (9)

[a] Structures of dermatan sulfate and heparan sulfate.

Dermatan sulfate



Heparan sulfate



[b] The numbers in parentheses refer to the enzymes that hydrolyze those bonds.

CLIN. CORR. 8.8 MUCOPOLYSACCHARIDOSES

A group of human genetic disorders characterized by excessive accumulation and excretion of the oligosaccharides of proteoglycans exists, collectively called mucopolysaccharidoses. These disorders result from a deficiency of one or more lysosomal hydrolases responsible for the degradation of dermatan and/or heparan sulfate. The enzymes lacking in specific mucopolysaccharidoses that have been identified are presented in Table 8.2. Presumably, for complete sequential degradation of the glycosaminoglycans, additional enzymes are required to hydrolyze bonds (4) *N*-acetyl β-galactosaminidase, (7) α-glucosaminidase, and (8) *N*-acetylglucosamine sulfatase, for which deficiency diseases remain unknown.

Although the chemical basis for this group of disorders is similar, their mode of inheritance as well as clinical manifestations may vary. Hurler and Sanfilippo are transmitted as autosomal recessives, while Hunter is sex-linked. Both Hurler and Hunter are characterized by skeletal abnormalities and mental retardation, which in severe cases may result in early death. In contrast, in the Sanfilippo syndrome, the physical defects are relatively mild, while the mental retardation is severe. Collectively, the incidence for all mucopolysaccharidoses is 1 per 30,000 births.

In addition to those listed in Table 8.2, some others are less well understood. Morquio syndrome involves impaired degradation of keratan sulfate, multiple sulfatase deficiency involves the deficiency of at least six sulfatases, and "I-cell" disease

involves a marked decrease in several hydrolase enzymes.

These disorders are amenable to prenatal diagnosis, since the pattern of metabolism by affected cells obtained from amniotic fluid is strikingly different from normal. Furthermore, a promising approach for correcting certain of these disorders is enzyme or gene replacement therapy.

Certain common structural features of the oligosaccharide-linked *N*-glycosidically to asparagine have also emerged. These glycoproteins commonly contain a core structure consisting of mannose (Man) residues linked to *N*-acetylglucosamine (GlcNAc) in the following structure:

$$(\text{Man})_n \rightarrow \text{Man} \xrightarrow[1,4]{} \text{GlcNAc} \xrightarrow[1,4]{} \text{GlcNAc} \rightarrow \text{Asn}$$

Synthesis of Glycoprotein

In contrast to the synthesis of *O*-glycosidically linked glycoproteins, which involves the sequential action of a series of glycosyltransferases, the synthesis of *N*-glycosidically linked peptides, involves a somewhat different and more complex mechanism (Figure 8.8). The common core is preassembled as a lipid-linked oligosaccharide prior to incorporation into the polypeptide. Similar assembly processes for synthesis of precursor units followed by transfer *en bloc* have been reported in bacterial cell wall synthesis, but are uncommon in mammalian heterosaccharide synthesis. During synthe-

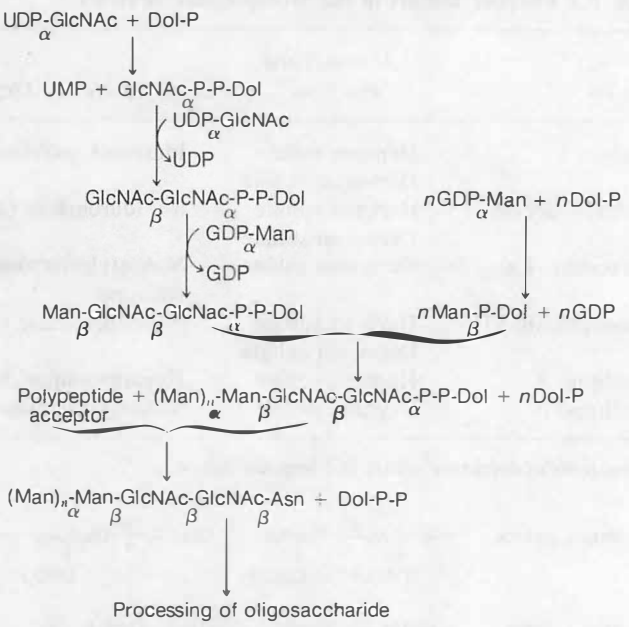
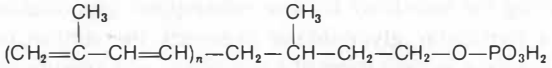


Figure 8.8
Biosynthesis of the oligosaccharide core in asparagine-N-acetylgalactosamine-linked glycoproteins.
Abbreviation: Dol, dolichol.

sis, the oligosaccharide intermediates are bound to derivatives of dolichol phosphate.



Dolichol phosphate

Dolichols are a class of polyprenols (C_{80}–C_{100}) containing 16 to 20 isoprene units, in which the final isoprene unit is saturated. These lipids participate in two types of reactions in core oligosaccharide synthesis. The first reaction involves formation of *N*-acetylglucosaminylpyrophosphoryldolichol with release of UMP from the respective nucleotide sugars. The second *N*-acetylglucosamine and mannose transferase reactions proceed by sugar transfer from the nucleotide without formation of intermediates. The subsequent addition of a mannose unit also occurs via a dolichol-linked mechanism. In the final step, the oligosaccharide is transferred from the dolichol pyrophosphate to an asparagine residue in the polypeptide chain.

After synthesis of the specific core region, the oligosaccharide chains are completed by the action of glycosyltransferases without further participation of lipid intermediates. Recent evidence suggests that extensive "processing," involving the addition and subsequent removal of certain glycosyl residues, occurs during the course of synthesis of asparagine-*N*-acetylglucosamine-linked glycoproteins. The pathways of "oligosaccharide processing" are currently under investigation in certain enveloped viruses (VSV and Sindbis) and in some secretory glycoprotein systems. It is thought that oligosaccharides destined to become simple or oligomannoside-type glycoproteins probably undergo more limited processing, whereas the more complex glycoproteins undergo more extensive processing and elongation. It is not yet understood why some glycoproteins become one type or another, but it may be of evolutionary significance.

Initiation of glycosylation by addition of the core saccharides may occur while the nascent peptide is still bound to ribosomes or soon after completion. Processing and elongation reactions then take place as the peptide moves through the rough endoplasmic reticulum to the Golgi. Once elongation is complete by addition of external sugars within the Golgi apparatus, the glycoprotein migrates toward the plasma membrane within a vesicle. The membrane of the transport vesicle fuses with the plasma membrane, and secretory glycoproteins are extruded, while internal membrane glycoproteins remain part of the plasma membrane.

Just as the synthesis of oligosaccharides requires specific glycosyltransferases, the degradation requires specific glycosidases. Exoglycosidases remove sugars sequentially from the nonreducing end, exposing the substrate for the subsequent glycosidase. The absence of a particular glycosidase prevents the action of the next enzyme, resulting in cessation of catabolism and accumulation of the product (Clin. Corr. 8.7 and Table 8.1). There is also evidence for the presence of endoglycosidases with broader specificity. However, the sequence of action of endo- and exoglycosidases in the catabolism of glycoproteins is not well understood. The primary degradation process occurs in lysosomes, but there are also specific microsomal glycosidases involved in the processing of glycoproteins during synthesis.

8.5 PROTEOGLYCANS

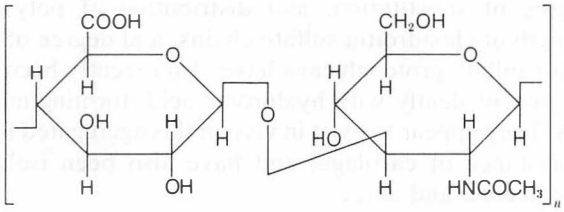
In addition to glycoproteins, which usually contain proportionally less carbohydrate than protein by weight, there is another class of complex macromolecules, containing 95% or more carbohydrate. The properties of these compounds resemble polysaccharides more than proteins. To distinguish these compounds from other glycoproteins, they are referred to as proteoglycans and their carbohydrate chains as glycosaminoglycans. An older name, mucopolysaccharide, is still in use, especially in reference to the group of storage diseases, mucopolysaccharidoses, which result from an inability to degrade these molecules.

The proteoglycans are high molecular weight polyanionic substances consisting of many different glycosaminoglycan chains linked covalently to a protein core. Although six distinct classes of glycosaminoglycans are now recognized, certain features are common to all classes. The long heteropolysaccharide chains are made up largely of disaccharide repeating units, in which one sugar is a hexosamine and the other a uronic acid. Other common constituents of glycosaminoglycans are sulfate groups, linked by ester bonds to certain monosaccharides or by amide bonds to the amino group of glucosamine. An exception, hyaluronic acid, is not sulfated, and as yet has not been shown to exist covalently attached to protein. The carboxyl and sulfate groups contribute to the nature of glycosaminoglycans as highly charged polyanions. Both their electrical charge and macromolecular structure aid in their biological role as lubricants and support elements in connective tissue. These

compounds form solutions with high viscosity and elasticity by absorbing large volumes of water. This allows them to act in stabilizing and supporting fibrous and cellular elements of tissues, as well as contributing to the maintenance of water and salt balance in the body.

Hyaluronic Acid

Among the glycosaminoglycans, hyaluronic acid is very different from the other five types. As previously mentioned, it is unsulfated, is not covalently complexed with protein, and is the only glycosaminoglycan not limited to animal tissue, but is also produced by bacteria. It is nevertheless classified as a glycosaminoglycan because of its structural similarity to these other polymers. It consists solely of repeating disaccharide units of *N*-acetylglucosamine and glucuronic acid. Although hyaluronic acid has the least complex



Repeat unit of hyaluronic acid

chemical structure of all the glycosaminoglycans, the chains may reach molecular weights of 10^5 to 10^7. The large molecular weight, polyelectrolyte character, and large volume of water it occupies in solution all contribute to the properties of hyaluronic acid as a lubricant and shock absorbant. Hence, it is found predominantly in synovial fluid, vitreous humor, and umbilical cord.

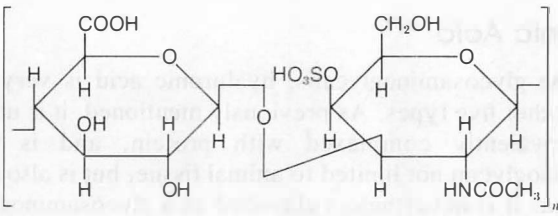
Chondroitin Sulfates

The most abundant glycosaminoglycans in the body are the chondroitin sulfates. The individual polysaccharide chains are attached to specific serine residues in a protein core of approximate molecular weight of 200,000 through a tetrasaccharide linkage region.

$$\text{GlcUA} \xrightarrow[1,3]{} \text{Gal} \xrightarrow[1,3]{} \text{Gal} \xrightarrow[1,4]{} \text{Xyl} \rightarrow \textit{O}\text{-Ser}$$

The characteristic repeating disaccharide units of *N*-acetylgalactosamine and glucuronic acid are covalently attached to this linkage region. The disaccharides may be sulfated in either

the 4 or 6 position of *N*-acetylgalactosamine. Each polysaccharide chain contains between 30 and 50 such disaccharide units, corresponding to molecular weights of 15,000 to 25,000. An average chondroitin sulfate proteoglycan molecule has approximately 100 chondroitin sulfate chains attached to the protein core, giving rise to

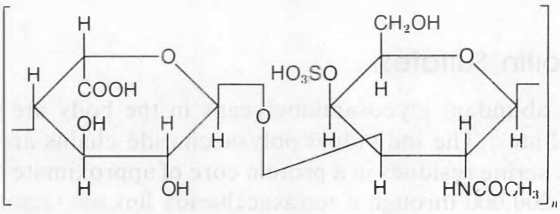


Repeat unit of chondroitin 4-sulfate

a molecular weight of 1.5 to 2 × 10^6. Proteoglycan preparations are, however, extremely heterogeneous, differing in length of protein core, degree of substitution, and distribution of polysaccharide chains, length of chondroitin sulfate chains, and degree of sulfation. Chondroitin sulfate proteoglycans have also recently been shown to aggregate noncovalently with hyaluronic acid, forming much larger structures. They appear to exist in vivo in this aggregated form in the ground substance of cartilage, and have also been isolated from tendon, ligaments, and aorta.

Dermatan Sulfate

Dermatan sulfate differs from chondroitin 4- and 6-sulfates in that its predominant uronic acid is L-iduronic acid, although D-glucuronic acid is also present in variable amounts. The glycosidic linkages are the same in position and configuration as in the chondroitin sulfates,



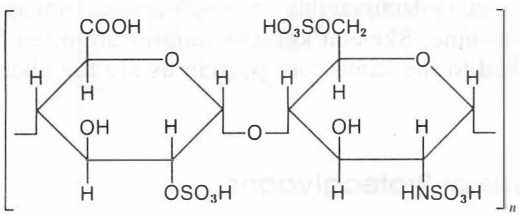
Repeat unit of dermatan sulfate

with average polysaccharide chains of molecular weights of 2 to 5 × 10^4. The physiological function of dermatan sulfate is poorly understood. Unlike the chondroitin sulfates, dermatan sulfate is antithrombic as is heparin, but in contrast to heparin, it shows only minimal whole blood anticoagulant and blood lipid-clearing ac-

tivities. As a connective tissue macromolecule, dermatan sulfate is found in skin, blood vessels, and heart valves.

Heparin and Heparan Sulfate

Heparin differs from other glycosaminoglycans in a number of important respects. Glucosamine and D-glucuronic acid or L-iduronic acid form the characteristic disaccharide repeat unit, as in dermatan sulfate. In contrast to the glycosaminoglycans in ground substance, heparin contains α-glycosidic linkages. Almost all glucosamine residues are bound in sulfamide linkages, but a small number of glucosamine residues are *N*-acetylated. The sulfate content of heparin, although variable, approaches 2.5 sulfate residues per disaccharide unit in preparations with the highest biological activity. In addition to *N*-sulfate and *O*-sulfate on C-6 of glucosamine, heparin may also contain sulfate on carbon-3 of the hexosamine and C-2 of



Repeat unit of heparin

the uronic acid. Unlike the other glycosaminoglycans, which are predominantly extracellular components, heparin is an intracellular component of mast cells. Heparin is known as an anticoagulant and lipid-clearing agent; however, the natural physiological role of this polysaccharide still remains unclear (see Clin. Corr. 8.9).

Heparan sulfate contains a similar disaccharide repeat unit but has more *N*-acetyl groups, fewer *N*-sulfate groups, and a lower degree of *O*-sulfate groups. Heparan sulfate appears to be extracellular in distribution and has been isolated from blood vessel walls, amyloid, and brain. Most recently, it has been shown to be an integral and ubiquitous component of the cell surface.

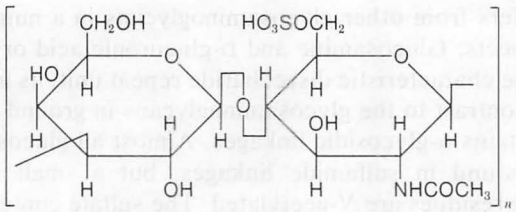
Keratan Sulfate

More than any of the other glycosaminoglycans, keratan sulfate is characterized by molecular heterogeneity. The polysaccharide is composed principally of a repeating disaccharide unit of *N*-acetylglucosamine and galactose, with no uronic acid in the molecule. Sulfate content is variable, with ester sulfate present on C-6 of

CLIN. CORR. 8.9 HEPARIN AS AN ANTICOAGULANT

Heparin is a naturally occurring sulfated polysaccharide that is used to reduce the clotting tendency of patients. Both in vivo and in vitro heparin prevents the activation of clotting factors, but does not act directly on the clotting factors. Rather, it is thought that the anticoagulant activity of heparin is brought about by the binding interaction of heparin with an inhibitor of the coagulation process. Presumably, heparin binding induces a conformational change in the inhibitor that generates a complementary interaction between the inhibitor and the activated coagulation factor, thereby preventing the factor from participating in the coagulation process. The inhibitor that interacts with heparin directly is antithrombin III, a plasma protein inhibitor of serine proteases. In the absence of heparin, antithrombin III slowly (10–30 min) combines with several clotting factors, yielding complexes devoid of proteolytic activity. In the presence of heparin, inactive complexes are formed within a few seconds.

both galactose and hexosamine. Two types of keratan sulfate have been distinguished, which differ in their overall carbohydrate content and tissue distribution. Both contain as additional monosaccharides, mannose, fucose, sialic acid, and *N*-acetylgalactosamine.



Repeat unit of keratan sulfate

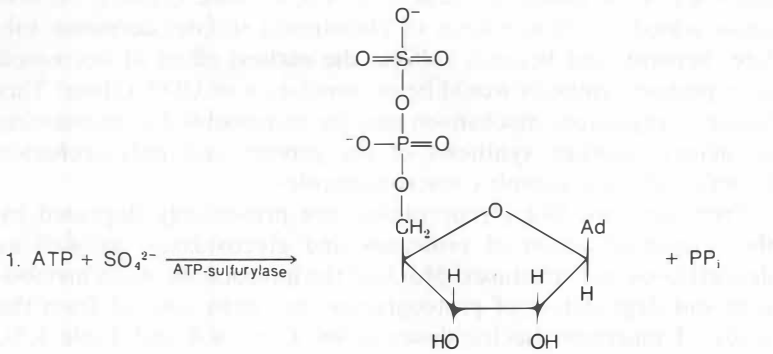
Keratan sulfate I, isolated from cornea, is linked to protein by an *N*-acetylglucosamine–asparaginyl bond, typical of glycoproteins. Keratan sulfate II, isolated from cartilage, is attached to protein through *N*-acetylgalactosamine in *O*-glycosidic linkage to either serine or threonine. Skeletal keratan sulfates are often found covalently attached to the same core protein as are the chondroitin sulfate chains.

Biosynthesis of Proteoglycans

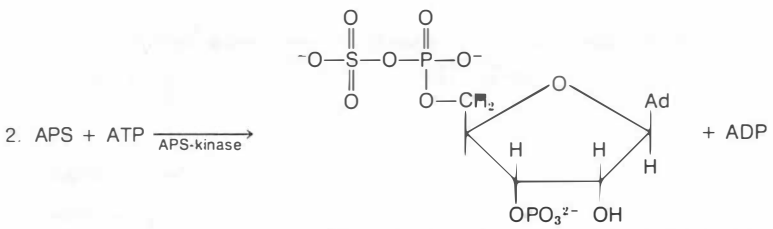
Most of the problems in studies on the biosynthesis of proteoglycans are similar to those previously encountered in the formation of other glycoproteins. The polysaccharide chains are assembled by the sequential action of a series of glycosyltransferases, which catalyze the transfer of a monosaccharide from a nucleotide sugar to an appropriate acceptor, either the nonreducing end of another sugar or a polypeptide. Since the biosynthesis of the chondroitin sulfates is most thoroughly understood, this pathway will be discussed as the prototype for glycosaminoglycan formation (Figure 8.9).

In analogy with the mechanisms established for glycoprotein biosynthesis, the formation of the core protein of the chondroitin sulfate proteoglycan is the first step in this process. However, whether the initiation of polysaccharide chains precedes completion of the peptide chain and release from the ribosomes has not been determined. The polysaccharide chains are assembled by six different glycosyltransferases. Strict substrate specificity is required for completion of the unique tetrasaccharide linkage region. Polymerization then results from the concerted action of two glycosyltransferases, an *N*-acetylgalactosaminyltransferase and a glucuronosyltransferase, which alternately add the two monosaccharides, forming the characteristic repeating disaccharide units.

Sulfation of *N*-acetylgalactosamine residues in either the 4 or 6 position apparently occurs along with chain elongation. The sulfate donor in these reactions, as in other biological systems, is 3′-phosphoadenosine 5′-phosphosulfate (PAPS), which is formed from ATP and sulfate in two steps.



Adenosine 5′-phosphosulfate (APS)



3′-Phosphoadenosine 5′-phosphosulfate (PAPS)

Synthesis of the other glycosaminoglycans requires additional transferases specific for the sugars and linkages found in these molecules. Completion of these glycosaminoglycans often involves modifications in addition to ester sulfation, including epimerization, acetylation, and *N*-sulfation. Interestingly, the epimerization of D-glucuronic acid to L-iduronic acid occurs after incorporation into the polymer chain and is coupled with the process of sulfation.

It still remains unclear what fundamental central mechanisms determine the quantity as well as the qualitative nature of the proteoglycans synthesized. It appears that different proteoglycans are synthesized because of the presence and strict substrate specificity of the enzymes and the formation of specific acceptor proteins in a cell. Two mechanisms by which the levels of nucleotide sugars may be regulated have previously been described. The first specific reaction in hexosamine synthesis, the fructose 6-phosphate : glutamine transamidase reaction (Figure 8.4), is subject to feedback inhibition by UDP-*N*-acetylglucosamine, which is in equilibrium with UDP-*N*-

acetylgalacosamine. Hence, synthesis of both proteoglycans and glycoproteins is regulated by the same mechanism. More specific to proteoglycan synthesis, the levels of UDP-xylose and UDP-glucuronic acid are stringently controlled by the inhibition by UDP-xylose of the UDP-glucose dehydrogenase conversion of UDP-glucose to UDP-glucuronic acid (Figure 8.4). Since xylose is the first sugar added during synthesis of chondroitin sulfate, dermatan sulfate, heparin, and heparan sulfate, the earliest effect of decreased core protein synthesis would be accumulation of UDP-xylose. This sensitive regulatory mechanism may be responsible for maintaining a balance between synthesis of the protein and polysaccharide moieties of these complex macromolecules.

Proteoglycans, like glycoproteins, are presumably degraded by the sequential action of proteases and glycosidases, as well as deacetylases and sulfatases. Much of the information about metabolism and degradation of proteoglycans has been derived from the study of mucopolysaccharidoses (Clin. Corr. 8.8 and Table 8.2).

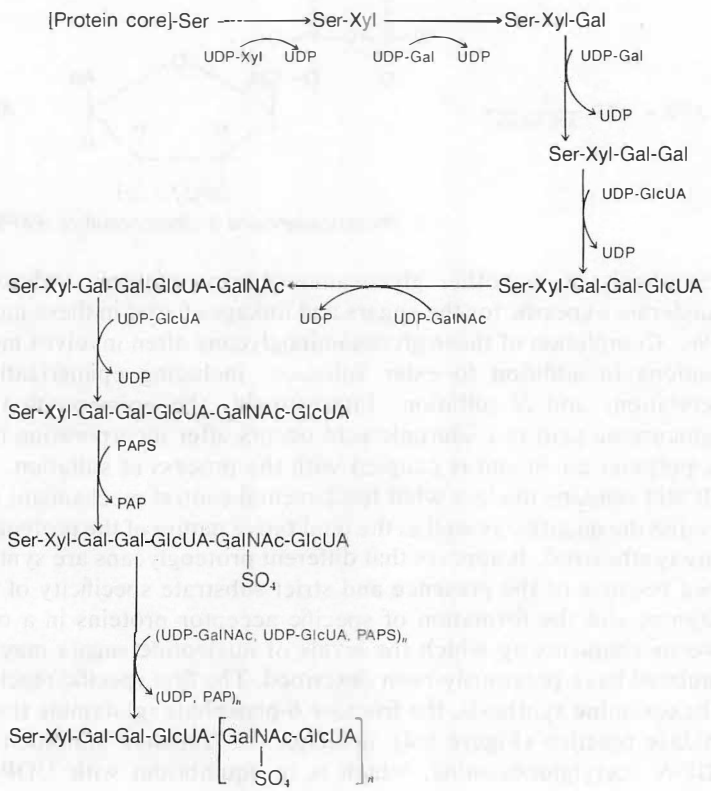


Figure 8.9
Synthesis of chondroitin sulfate proteoglycan. *Abbreviations: Xyl, xylose; Gal, galactose; GlcUA, glucuronic acid; GalNAc,* N*-acetylgalactosamine; PAPS, phosphoadenosine phosphosulfate.*

This group of human genetic disorders is characterized by accumulation in tissues and excretion in urine of oligosaccharide products derived from incomplete breakdown of the proteoglycans, due to a deficiency of one or more lysosomal hydrolases. In the diseases for which the biochemical defect has been identified, it has been shown that a product accumulates with a nonreducing terminus that would have been the substrate for the deficient enzyme.

BIBLIOGRAPHY

Dutton, G. J. (ed.). *Glucuronic acid, free and combined*. New York: Academic, 1966.

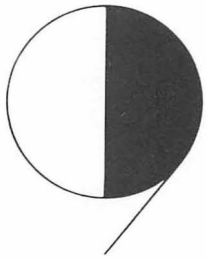
Horecker, B. L. *Pentose metabolism in bacteria*. New York: Wiley, 1962.

Margolis, R. V., and Margolis, R. K. (eds.). *Complex carbohydrates of nervous tissue*. New York: Plenum, 1979.

Sharon, N. *Complex carbohydrates—their chemistry, biosynthesis and functions*. Reading, Mass.: Addison-Wesley, 1975.

Stanbury, J. B., Wyngarden, J. B., and Frederickson, D. S. (eds.). *The metabolic basis of inherited disease*. 3rd ed. New York: McGraw-Hill, 1972.

Walborg, E. F., Jr. (ed.). Glycoproteins and glycolipids in disease processes, in ACS Symposium Series 80. Washington, D.C.: American Chemical Society, 1978.



Lipid Metabolism I

FRANCIS N. LeBARON

UTILIZATION AND STORAGE OF ENERGY IN LIPID FORM

9.1 OVERVIEW

As the human body builds and renews its structures, obtains and stores energy, and performs its various functions, there are numerous circumstances in which it is essential to use molecules or parts of molecules that do not associate with water. This property of being nonpolar and hydrophobic is largely supplied by the substances classed as lipids. Most of these are molecules that contain or are derived from fatty acids. In the early stages of biochemical research these substances were not studied as intensively as other body constituents, largely because the techniques for studying aqueous systems were easier to develop. This benign neglect led to early assumptions that the lipids were relatively inert and their metabolism was of lesser importance than that of carbohydrates, for instance.

As the methodology for investigation of lipid metabolism developed, however, it soon became evident that fatty acids and their derivatives had at least two major roles in the human body. On the one hand, the oxidation of fatty acids was shown to be a major means of metabolic energy production, and it became clear that their storage in the form of triacylglycerols was more efficient and quantitatively more important than storage of carbohydrates as glycogen. On the other hand, as details of the chemistry of biological structures were elucidated, hydrophobic structures were found to be largely composed of fatty acids and their derivatives. Thus the major separation of cells and subcellular structures into separate aqueous compartments is accomplished by the use of membranes whose hydrophobic characteristics are largely supplied by the fatty acid moieties of complex lipids. These latter compounds contain constituents other than fatty acids and glycerol. They frequently have significant covalently-bound hydrophilic moieties, notably carbohydrates in the glycolipids and organic phosphate esters in the phospholipids.

In addition to these two major functions of lipids, energy production and structure building, there are several other quantitatively less important roles, which are nonetheless of great functional significance. These include the use of the surface active properties of some complex lipids for specific functions, such as maintenance of lung alveolar integrity and solubilization of nonpolar substances in body fluids. In addition, several classes of lipids, the steroid hormones and the prostaglandins, have highly potent and specific physiological roles in control of metabolic processes. The interrelationships of all the processes involved in lipid metabolism are outlined in Figure 9.1.

Since the metabolism of fatty acids and triacylglycerols is so crucial to proper functioning of the human body, imbalances and

deficiencies in these processes lead to significant pathological processes, and disease states related to fatty acid and triacylglycerol metabolism include some of the major clinical problems to be encountered by physicians, for instance, ketoacidosis, obesity, and errors in transport of lipids in blood. In addition, some unique genetic deficiencies have been found, such as Refsum's disease, which have helped to elucidate some pathways in lipid metabolism.

In this first chapter on lipid metabolism we will be primarily concerned with the structure and metabolism of the fatty acids themselves and of their primary storage form, the triacylglycerols. After an initial discussion of the structures of the more important fatty acids in the human body we describe how they are supplied to the human metabolic machinery from the diet or by biosynthesis. Since their storage as triacylglycerols is a major process we next discuss how this storage is accomplished and how the fatty acids themselves

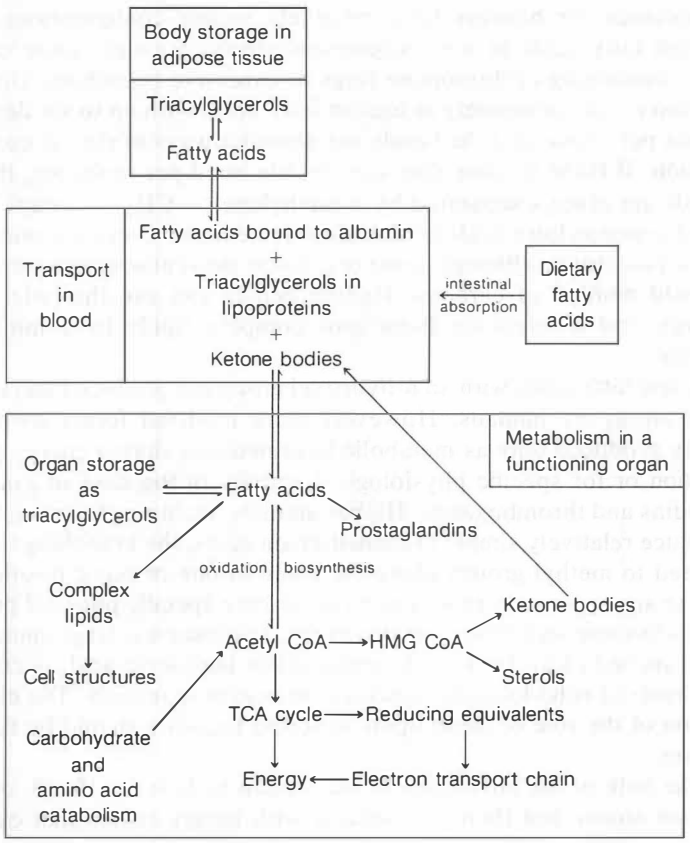


Figure 9.1
The metabolic interrelationships of fatty acids in the human body.

are mobilized and transported throughout the body to sites where they are needed. The central process of energy production from fatty acids is then discussed, and finally we introduce some concepts about the role and metabolism of polyunsaturated fatty acids.

9.2 THE CHEMICAL NATURE OF FATTY ACIDS AND ACYLGLYCEROLS

The Structure of Fatty Acids

Fatty acids consist of an alkyl chain with a terminal carboxyl group, and the simplest configuration is a completely saturated straight chain. The basic formula is $CH_3—(CH_2)_n—COOH$. The fatty acids of importance for humans have relatively simple configurations, although fatty acids in some organisms are occasionally quite complex, containing cyclopropane rings or extensive branching. Unsaturation occurs commonly in human fatty acids with up to six double bonds per chain, and the bonds are almost always of the cis configuration. If there is more than one double bond per molecule, these bonds are always separated by a methylene ($—CH_2—$) group. The most common fatty acids in biological systems have an even number of carbon atoms, although some organisms do synthesize those with an odd number of carbons. Human beings can use the latter for energy and incorporate them into complex lipids to a minimal degree.

A few fatty acids with an α-hydroxyl group are produced and used structurally by humans. However, more oxidized forms are normally produced only as metabolic intermediates during energy production or for specific physiological activity in the case of prostaglandins and thromboxanes. Higher animals, including humans, also produce relatively simple branched-chain acids, the branching being limited to methyl groups along the chain at one or more positions. These are apparently produced to contribute specific physical properties to some secretions and structures. For instance, large amounts of branched chain fatty acids, particularly isovaleric acid, occur in the lipids of echo-locating structures in marine mammals. The elucidation of the role of these lipids in sound focusing should be fascinating.

Isovaleric acid

The bulk of the fatty acids in the human body have 16, 18, or 20 carbon atoms, but there are several with longer chains that occur

principally in the lipids of the nervous system. These include nervonic acid, and a 22-carbon acid with six double bonds.

$CH_3—(CH_2)_7—CH=CH—(CH_2)_{13}—COOH$

Nervonic acid

$CH_3—(CH_2—CH=CH)_6—(CH_2)_2—COOH$

All-*cis*-4,7,10,13,16,19-docosahexaenoic acid

Fatty Acid Nomenclature

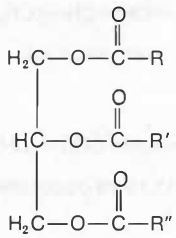
The most abundant fatty acids have common names that have been accepted for use in the official nomenclature. Examples are given in Table 9.1 along with the official systematic names. The approved abbreviations consist of the number of carbon atoms followed, after a colon, by the number of double bonds. Carbon atoms are numbered with the carboxyl carbon as number 1, and the double bond locations are designated by the number of the carbon atom on the carboxyl side of it. These designations of double bonds are in parentheses after the rest of the symbol. See Table 9.1 for examples.

Table 9.1 Fatty Acids of Importance to Humans

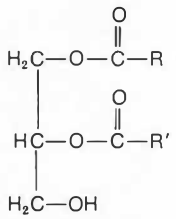
Numerical Symbol	*Structure*	*Trivial Name*	*Systematic Name*
16:0	$CH_3—(CH_2)_{14}—COOH$	Palmitic	Hexadecanoic
16:1(9)	$CH_3—(CH_2)_5—CH=CH—(CH_2)_7—COOH$	Palmitoleic	*cis*-9-Hexadecenoic
18:0	$CH_3—(CH_2)_{16}—COOH$	Stearic	Octadecanoic
18:1(9)	$CH_3—(CH_2)_7—CH=CH—(CH_2)_7—COOH$	Oleic	*cis*-9-Octadecenoic
18:2(9,12)	$CH_3—(CH_2)_3—(CH_2—CH=CH)_2—(CH_2)_7—COOH$	Linoleic	*cis,cis*-9,12-Octadecadienoic
18:3(9,12,15)	$CH_3—(CH_2—CH=CH)_3—(CH_2)_7—COOH$	Linolenic	*cis,cis,cis*-9,12,15-Octadecatrienoic
20:4(5,8,11,14)	$CH_3—(CH_2)_3—(CH_2—CH=CH)_4—(CH_2)_3—COOH$	Arachidonic	*cis,cis,cis,cis*-5,8,11,14-Icosatetraenoic

Structure of Acylglycerols

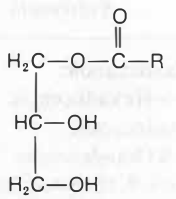
Fatty acids occur primarily as esters of glycerol when they are being used for energy storage and utilization. Compounds with one (monoacylglycerols) or two (diacylglycerols) acids esterified are present only in relatively minor amounts and occur largely as metabolic intermediates in the biosynthesis and degradation of glycerol-containing lipids. The bulk of the fatty acids in the human body exist as triacylglycerols, in which all three hydroxyl groups on the glycerol are esterified with a fatty acid. Historically, these compounds have been termed *neutral fats* or *triglycerides,* and these terms are still in common usage. However, there are other types of "neutral fats" in the body, and the term "triglyceride" is chemically incorrect and should no longer be used. The same can be said for the terms "monoglyceride" and "diglyceride."



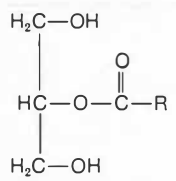
Triacylglycerol



Diacylglycerol



1-Monoacylglycerol



2-Monoacylglycerol

The distribution of various fatty acids in the different positions of the triacylglycerols in the body at any given time is the result of a number of factors, some of which are not completely understood. Suffice it to say that the fatty acid pattern varies with the time, diet, and anatomical location of the triacylglycerol. Compounds with the same fatty acid in all three positions are rare and the usual case is for a complex mixture.

Physical and Chemical Properties

Certainly one of the most prominent and significant properties of fatty acids and triacylglycerols is their lack of affinity for water. The long hydrocarbon chains have negligible possibility for hydrogen bonding, and the acids, whether unesterified or in a complex lipid, have a much greater tendency to associate with each other or with other hydrophobic structures, such as sterols and the hydrophobic side chains of amino acids, than they do with water or polar organic compounds. It has been calculated that the van der Waals–London forces between closely packed, relatively long-chain fatty acid moieties in lipids can approach the strength of a covalent bond. This hydrophobic character is essential for construction of complex biological structures and the separation of aqueous compartments as described in Chapter 5. It is also essential for use in biological surface active molecules, as in the intestinal tract.

Of major significance in the context of this chapter is the fact that the hydrophobic nature of triacylglycerols and their relatively reduced state make them efficient compounds for storing energy. It can be estimated that on a weight basis pure triacylglycerols yield nearly $2\frac{1}{2}$ times the amount of ATP on complete oxidation that pure glycogen does. In addition, the triacylglycerols can be stored as pure lipid without associated water, whereas glycogen is quite hydrophilic and binds about twice its weight of water when stored in tissues. Thus the equivalent amount of metabolically recoverable energy stored as hydrated glycogen would weigh about four times as much as if it were stored as triacylglycerols.

The bulk of the fatty acids in the lipids of the human body are either saturated or contain only one double bond. Consequently, although they are readily catabolized by appropriate enzymes and cofactors, they are fairly inert chemically. This is an added advantage of their use for energy storage. However, the smaller amounts of the more highly unsaturated fatty acids in the tissues are much less stable to oxidation. The possible biological consequences of this oxidation are discussed later in this chapter.

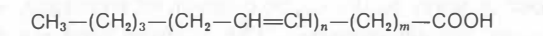
9.3 SOURCES OF FATTY ACIDS

Both diet and biosynthesis supply the fatty acids needed by the human body for energy and for construction of hydrophobic parts of biomolecules. Excess amounts of protein and carbohydrate obtained in the diet are readily converted to fatty acids and stored as triacylglycerols.

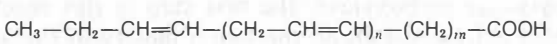
Dietary Supply

A great proportion of the fatty acids utilized by humans is supplied in their diet. Various animal and vegetable lipids are ingested, hydrolyzed at least partially by digestive enzymes, and absorbed through the intestinal wall to be distributed through the body, first in the lymphatic system and then in the bloodstream. These processes are extensively discussed in Chapter 24. To some extent, then, dietary supply governs the composition of the fatty acids in the body lipids. On the other hand, metabolic processes in the tissues of the normal human body can modify the dietary fatty acids, and/or those that are synthesized in these tissues, to produce almost all the various structures that are needed. For this reason, with one exception, the actual composition of the fatty acids supplied in the diet is relatively unimportant. This one exception involves the need for appropriate proportions of the relatively highly unsaturated fatty acids and particularly relates to the fact that many higher mammals, including humans, are unable to produce fatty acids with double bonds very far toward the methyl end of the molecule, either during de novo synthesis or by modification of dietary acids. Despite this inability, certain polyunsaturated acids with double bonds within the last seven linkages toward the methyl end are essential for some specific functions. Although all the reasons for this need are not yet elucidated, certainly one is that some of these acids are precursors of prostaglandins, highly active oxidation products (see Chapter 10).

In humans a dietary precursor is essential for two series of fatty acids. These are the linoleic series and the linolenic series.



Basic formula of the linoleic acid series



Basic formula of the linolenic acid series

Biosynthesis of Palmitate, the "Stem" Fatty Acid

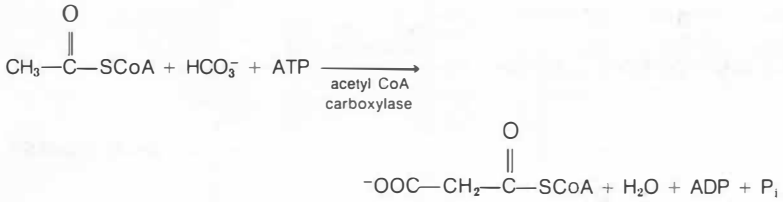
Besides dietary supply, the second major source of fatty acids for humans is their biosynthesis from small-molecule intermediates, which can be derived from metabolic breakdown of sugars, of some amino acids, and of other fatty acids. In a majority of instances the saturated, straight-chain 16-carbon acid, palmitic acid, is first synthesized, and all other fatty acids are made by modifications of palmitic acid. Acetyl CoA is the ultimate source of all carbon atoms for this synthesis, and the fatty acids are made by sequential addition of two-carbon units to the activated carboxyl end of the growing chain. In mammalian systems the sequence of reactions is carried out by fatty acid synthetase.

Fatty acid synthetase is a fascinating enzyme complex which is still being intensively studied. It is either an association of a few proteins, each of which has multiple enzyme functions, or it is a very tightly bound group of enzymes that have not yet been separated. For the most part its function is to form palmitate, but in some circumstances this pathway can be altered to produce other short-chain fatty acids. Some of the details of these modifications are discussed in later paragraphs, but first we will outline the basic scheme for synthesis of palmitate.

Either acetyl CoA or butyryl CoA is the priming unit for fatty acid synthesis, and the methyl end of these primers becomes the methyl end of palmitate. The addition of the rest of the two-carbon units requires further activation of the methyl carbon of acetyl CoA by carboxylation to malonyl CoA. However, the CO_2 added in this process is lost when the condensation of malonyl CoA to the growing chain occurs, so the carbons in the palmitate chain originate only from the acetyl CoA.

Formation of Malonyl CoA

The metabolic process that commits acetyl CoA to fatty acid synthesis rather than to its other many fates is its carboxylation to malonyl CoA by the enzyme acetyl CoA carboxylase. This reaction is similar in a number of ways to the carboxylation of pyruvate, which starts the process of gluconeogenesis. The reaction requires energy from ATP and uses dissolved bicarbonate as the source of CO_2. As in the case of pyruvate carboxylase, the first step in this reaction is the formation of activated CO_2 on the biotin moiety of the acetyl CoA carboxylase enzyme using the energy from ATP. This is then transferred to the acetyl CoA.



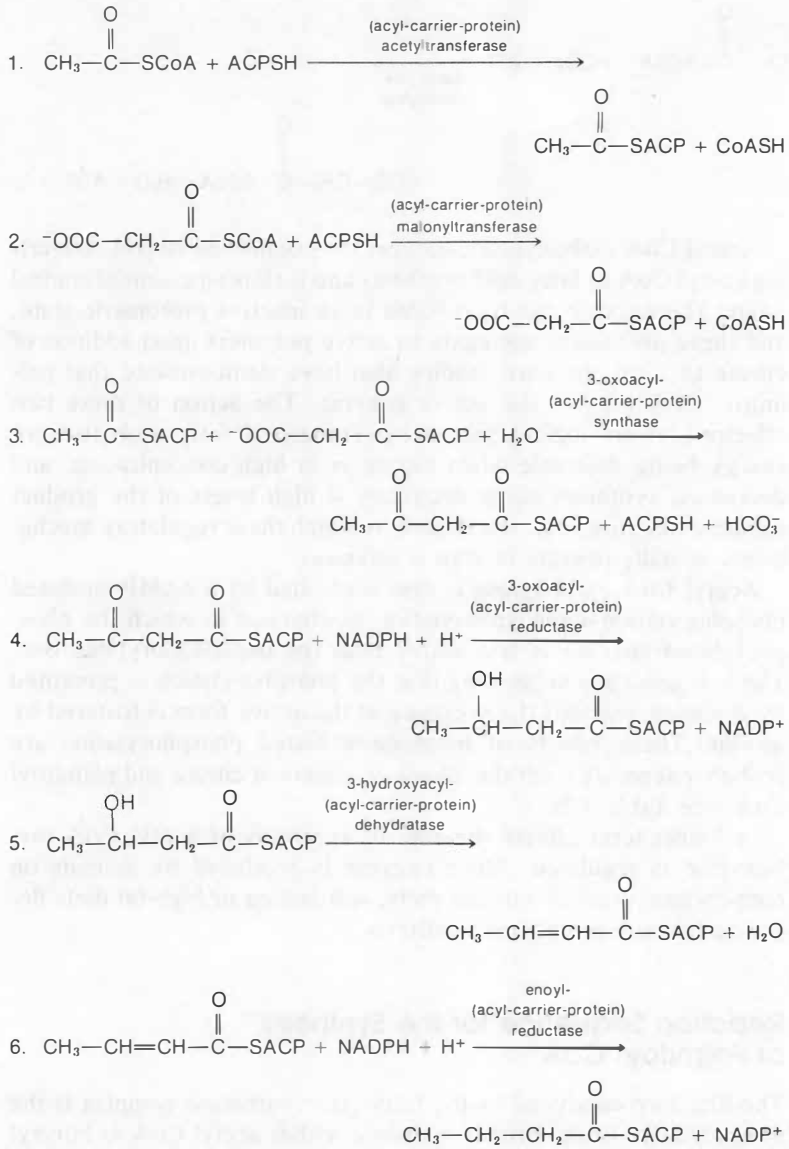
Acetyl CoA carboxylase catalyzes the committed step in converting acetyl CoA to fatty acid synthesis and is thus an essential control point. The enzyme can be isolated in an inactive protomeric state, and these protomers aggregate to active polymers upon addition of citrate in vitro. In vitro studies also have demonstrated that palmitoyl CoA inhibits the active enzyme. The action of these two effectors is very logical; increased synthesis of fatty acids to store energy being desirable when citrate is in high concentration, and decreased synthesis being necessary if high levels of the product accumulate. However, the degree to which these regulatory mechanisms actually operate in vivo is unknown.

Acetyl CoA carboxylase is also controlled by a cAMP-mediated phosphorylation–dephosphorylation mechanism in which the phosphorylated enzyme is less active than the dephosphorylated one. There is evidence suggesting that the phosphorylation is promoted by glucagon and that the presence of the active form is fostered by insulin. These effects of hormone-mediated phosphorylation are probably separate from the allosteric effects of citrate and palmitoyl CoA (see Table 9.2).

In longer-term effects the rate of synthesis of acetyl CoA carboxylase is regulated. More enzyme is produced by animals on high-carbohydrate or fat-free diets, and fasting or high-fat diets decrease the rate of enzyme synthesis.

Reaction Sequence for the Synthesis of Palmitoyl CoA

The first step catalyzed by the fatty acid synthetase complex is the transacylation of the primer molecule, either acetyl CoA or butyryl CoA, to a 4′-phosphopantetheine moiety on a protein constituent of the enzyme complex. This protein is acyl carrier protein (ACP), and its phosphopantetheine unit is identical to that in coenzyme A. Six or seven two-carbon units are then sequentially added to the enzyme complex until the palmitate molecule is completed. After each addition of a two-carbon unit a series of reductive steps takes place. The reaction sequence starting with an acetyl CoA primer is as follows:



The next sequence is initiated by addition of another malonyl ACP:

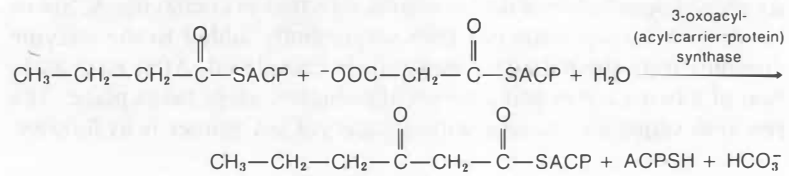
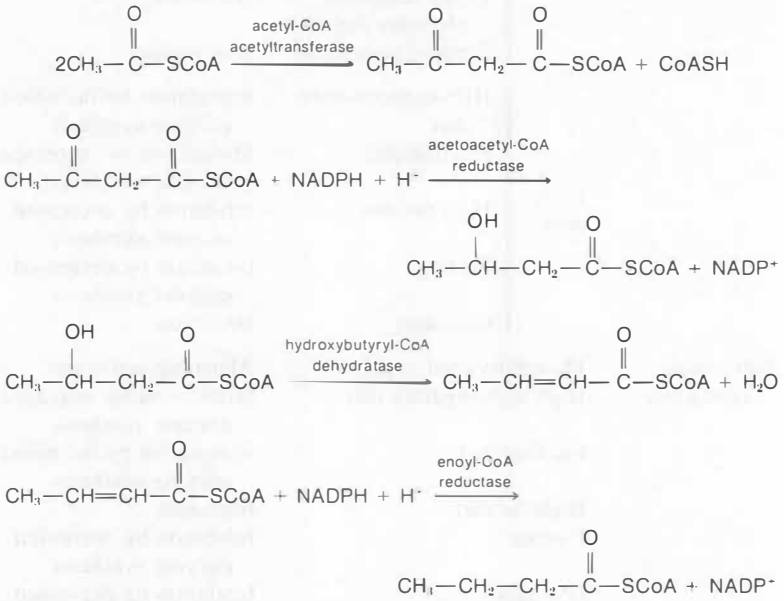


Table 9.2 Regulation of Fatty Acid Synthesis

Enzyme	*Regulatory Agent*		*Effect*
	PALMITATE BIOSYNTHESIS		
Acetyl CoA carboxylase	Short term	Citrate	Allosteric activation
		C_{16}–C_{18} acyl CoAs	Allosteric inhibition
		Insulin	Stimulation
		Glucagon	Inhibition
		cAMP-mediated phosphorylation	Inhibition
		Dephosphorylation	Stimulation
	Long term	High-carbohydrate diet	Stimulation by increased enzyme synthesis
		Fat-free diet	Stimulation by increased enzyme synthesis
		High-fat diet	Inhibition by decreased enzyme synthesis
		Fasting	Inhibition by decreased enzyme synthesis
		Glucagon	Inhibition
Fatty acid synthetase	Phosphorylated sugars		Allosteric activation
	High carbohydrate diet		Stimulation by increased enzyme synthesis
	Fat-free diet		Stimulation by increased enzyme synthesis
	High-fat diet		Inhibition
	Fasting		Inhibition by decreased enzyme synthesis
	Glucagon		Inhibition by decreased enzyme synthesis
	BIOSYNTHESIS OF FATTY ACIDS OTHER THAN PALMITATE		
Fatty acid synthetase	High ratio of $\frac{\text{methylmalonyl CoA}}{\text{malonyl CoA}}$		Increased synthesis of methylated fatty acids
	Thioesterase cofactor		Termination of synthesis with short-chain product
Stearoyl CoA desaturase	Various hormones		Stimulation of unsaturated fatty acid synthesis by increased enzyme synthesis
	Dietary polyunsaturated fatty acids		Decreased activity

The reductive steps then take place again exactly as in the reduction of acetoacetyl ACP above (reactions 4 to 6). Five more malonyl ACP then add to the growing chain by repetition of steps 3 to 6 and the product is palmitoyl ACP. The final step is hydrolysis of the palmitate from ACP, yielding free palmitic acid.

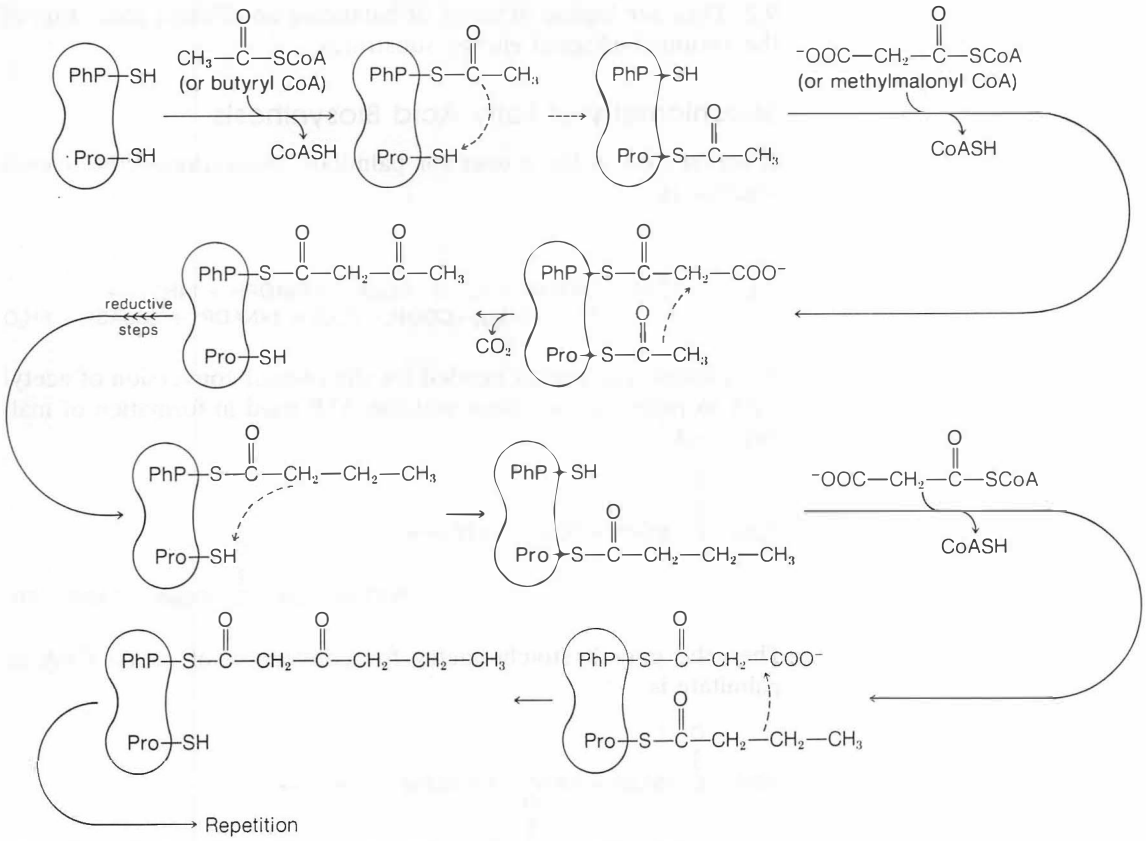
When the primer for fatty acid synthetase is butyryl CoA, acetyl CoA is still the ultimate source of the carbons. Butyryl CoA is formed from acetyl CoA by the action of cytoplasmic enzymes as follows:



Mammalian Fatty Acid Synthetase

The reaction sequence given above is fairly well established as the basic pattern for fatty acid biosynthesis in living systems. However, the details of the reaction mechanisms are still far from clear and may vary among species. The enzyme complex termed fatty acid synthetase catalyzes all these reactions, but its structure and properties vary considerably. The enzymes in *E. coli* are dissociable, and the reaction sequence was worked out with that organism. This sequence has been confirmed in mammalian systems, but the enzyme complex itself has not been dissociated. Some investigators postulate that the mammalian synthetase is composed of two possibly identical subunits, each of which is a multienzyme polypeptide. Even among mammalian species and tissues there are certainly variations.

Despite the gaps in present knowledge it appears likely that the growing fatty acid chain is continually bound to the enzyme complex and may be sequentially transferred between one or two 4′-phosphopantetheine groups and the sulfhydryl group of an enzyme-bound cysteine during the condensation reaction (reaction 3) (see Figure 9.2). It is also probable that an intermediate acylation to an enzyme serine takes place when acyl CoA units add to the enzyme-bound ACP in the transacylase reactions.



PhP = 4′-phosphopantetheine on acyl carrier protein
Pro = enzyme protein cysteine

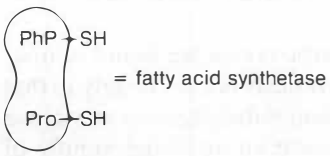


Figure 9.2
Proposed mechanism of elongation reactions taking place on mammalian fatty acid synthetase.

There is suggestive evidence that some short-term regulation of the fatty acid production is carried out by control of the activity of fatty acid synthetase, but this is yet to be established firmly. An allosteric stimulation by phosphorylated sugars has been proposed and a reversible control mechanism by addition or removal of the 4′-phosphopantetheine cofactor is suggested. However, most control of palmitate biosynthesis through fatty acid synthetase probably occurs by controlling the rate of synthesis and degradation of the enzyme. The agents and conditions which do this are given in Table 9.2. They are logical in terms of balancing an efficient utilization of the various biological energy substrates.

Stoichiometry of Fatty Acid Biosynthesis

If acetyl CoA is the primer for palmitate biosynthesis, the overall reaction is

$$CH_3{-}\overset{\overset{O}{\|}}{C}{-}SCoA + 7HOOC{-}CH_2{-}\overset{\overset{O}{\|}}{C}{-}SCoA + 14NADPH + 14H^+ \longrightarrow$$
$$CH_3{-}(CH_2)_{14}{-}COOH + 7CO_2 + 14NADP^+ + 8CoASH + 6H_2O$$

To calculate the energy needed for the overall conversion of acetyl CoA to palmitate, we must add the ATP used in formation of malonyl CoA:

$$7CH_3{-}\overset{\overset{O}{\|}}{C}{-}SCoA + 7CO_2 + 7ATP \longrightarrow$$
$$7HOOC{-}CH_2{-}\overset{\overset{O}{\|}}{C}{-}SCoA + 7ADP + 7P_i$$

Then the overall stoichiometry for conversion of acetyl CoA to palmitate is

$$8CH_3{-}\overset{\overset{O}{\|}}{C}{-}SCoA + 7ATP + 14NADPH + 14H^+ \longrightarrow$$
$$CH_3{-}(CH_2)_{14}{-}\overset{\overset{O}{\|}}{C}{-}OH + 8CoASH + 7ADP + 7P_i + 6H_2O + 14NADP^+$$

Subcellular Localization and Sources of Substrates

Fatty acid synthetase and acetyl CoA carboxylase are found primarily in the cytoplasm, and palmitate biosynthesis occurs largely in that subcellular compartment. However, mammalian tissues must use special processes to some extent to ensure an adequate supply of acetyl CoA and NADPH for this synthesis.

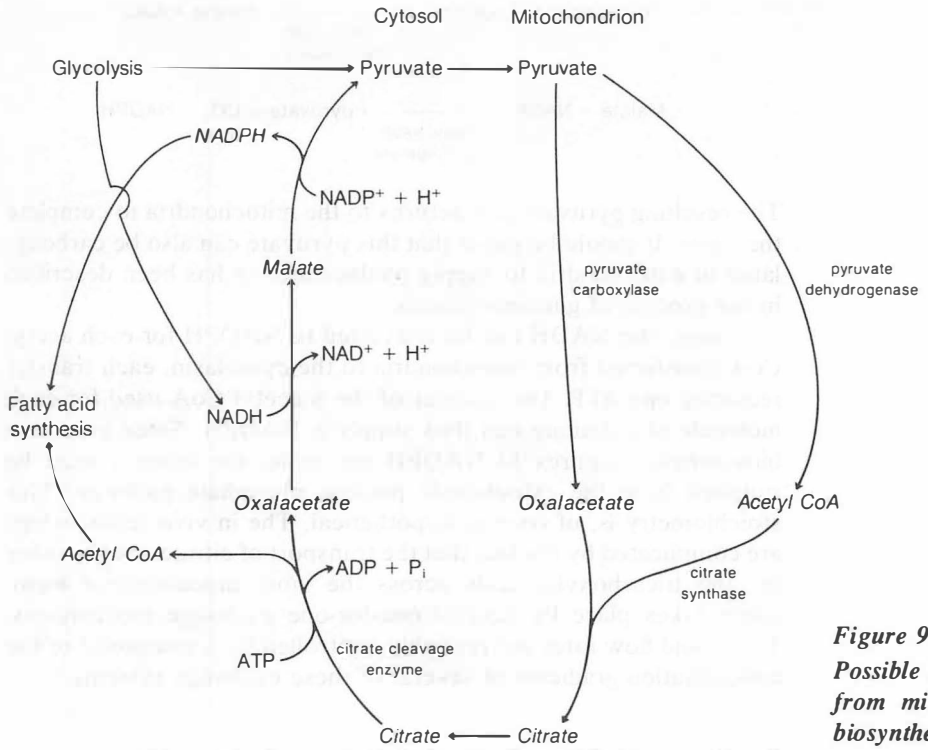


Figure 9.3
Possible mechanism for transfer of acetyl CoA from mitochondria to cytosol for fatty acid biosynthesis.

Specifically, the major source of acetyl CoA is the pyruvate dehydrogenase reaction inside the mitochondria. Since mitochondria are not readily permeable to acetyl CoA, a bypass mechanism moves it to the cytoplasm for palmitate biosynthesis. This mechanism, outlined in Figure 9.3, takes advantage of the facts that citrate does exchange freely from the mitochondria to the cytoplasm and that an enzyme exists in the cytoplasm to convert citrate to acetyl CoA and oxalacetate. When there is an excess of citrate for the tricarboxylic acid cycle, citrate will pass into the cytoplasm and supply acetyl CoA for fatty acid biosynthesis. Citrate cleavage enzyme catalyzes the cleavage which requires a mole of ATP:

$$\text{Citrate} + \text{ATP} + \text{CoA} \xrightarrow[\text{enzyme}]{\text{citrate cleavage}} \text{acetyl CoA} + \text{ADP} + P_i + \text{oxalacetate}$$

This mechanism has other advantages because CO_2 and NADPH for synthesis of palmitate can be produced from excess cytoplasmic oxalacetate. As shown in Figure 9.3, the process produces NADPH from NADH, which was formed during glycolysis, and two different malate dehydrogenases convert oxalacetate first to malate and then to pyruvate and CO_2.

$$\text{Oxalacetate} + \text{NADH} + \text{H}^+ \xrightarrow[\text{dehydrogenase}]{\text{malate-NAD}} \text{malate} + \text{NAD}^+$$

$$\text{Malate} + \text{NADP}^+ \xrightarrow[\text{dehydrogenase}]{\text{malate-NADP}} \text{pyruvate} + \text{CO}_2 + \text{NADPH}$$

The resulting pyruvate now returns to the mitochondria to complete the cycle. It should be noted that this pyruvate can also be carboxylated in mitochondria to supply oxalacetate, as has been described in the process of gluconeogenesis.

In sum, one NADH can be converted to NADPH for each acetyl CoA transferred from mitochondria to the cytoplasm, each transfer requiring one ATP. The transfer of the 8 acetyl CoA used for each molecule of palmitate can thus supply 8 NADPH. Since palmitate biosynthesis requires 14 NADPH per mole, the other 6 must be supplied from the cytoplasmic pentose phosphate pathway. This stoichiometry is, of course, hypothetical. The in vivo relationships are complicated by the fact that the transport of citrate and the other di- and tricarboxylic acids across the inner mitochondrial membrane takes place by several one-for-one exchange mechanisms. The actual flow rates are probably controlled by a composite of the concentration gradients of several of these exchange systems.

Synthesis of Other Fatty Acids from Palmitate

The human body can synthesize all of the fatty acids it needs except for the essential, polyunsaturated fatty acids. These syntheses involve a variety of enzyme systems in a number of locations, and the palmitic acid produced by fatty acid synthetase is modified by three processes: elongation, desaturation, and hydroxylation. There are probably two types of hydroxylation, one involved in introduction of an α-hydroxyl group on certain long-chain fatty acids required for synthesis of sphingolipids in the nervous system and the other involved as a step in the oxidation of fatty acids. In this section we discuss the process of elongation, the initial mechanism of desaturation and the hydroxylation of the brain fatty acids. The other α-oxidation process is considered later in the discussion of fatty acid oxidation, and the more involved desaturation schemes producing polyunsaturated fatty acids are outlined in a later section.

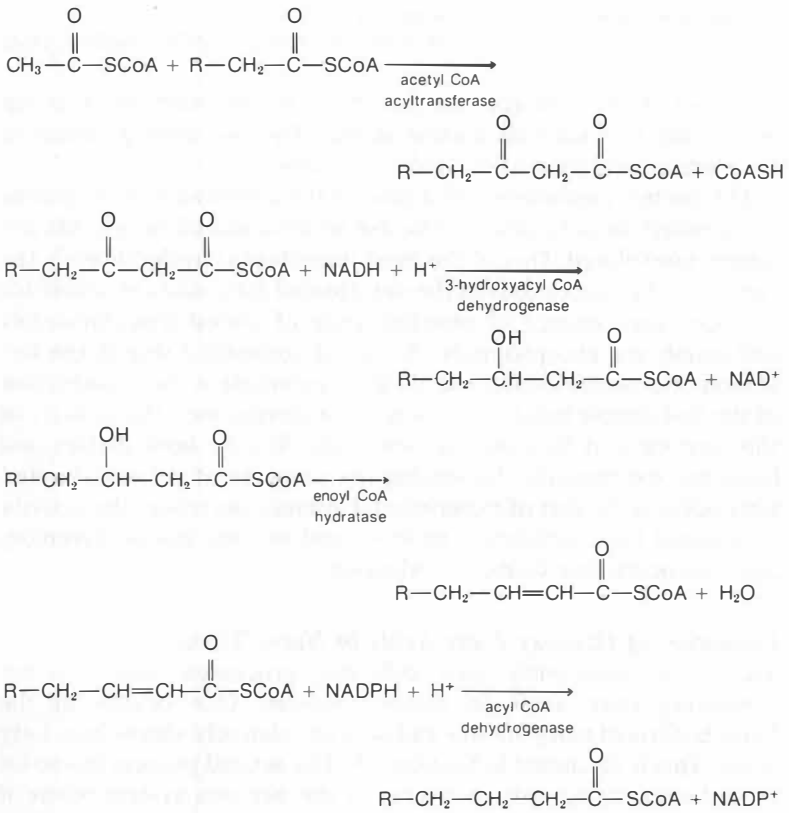
Elongation Reactions

In mammalian systems elongation of palmitate can occur either in the endoplasmic reticulum or in mitochondria, and the processes are slightly different in these two loci. In the endoplasmic reticulum the sequence of reactions is the same as that which occurs in the cyto-

plasmic fatty acid synthetase and the source of two-carbon units is malonyl CoA. The preferred substrate for elongation in most cases is palmitoyl CoA. The rate-controlling step appears to be initial condensation with malonyl CoA, and the intermediates in subsequent reactions are CoA esters, suggesting that the process is carried out by separate enzymes rather than a complex of the fatty acid synthetase type.

Present evidence suggests that in most tissues this elongation system in the endoplasmic reticulum almost exclusively converts palmitate to stearate. However, brain contains one or more additional elongation systems, which synthesize the longer chain acids (up to 24 C) that are needed for the brain lipids. These other systems also use malonyl CoA as substrate.

The elongation system in mitochondria is different from that in the endoplasmic reticulum in that acetyl CoA is the source of added two-carbon units. The reactions are as follows:



This process produces several fatty acids with longer chains than palmitic acid by using enzymes with differing substrate specificities.

Formation of Monoenoic Acids by Stearoyl CoA Desaturase

In higher animals desaturation of fatty acids occurs in the endoplasmic reticulum, and the oxidizing system used to introduce cis double bonds is significantly different from the main fuel oxidation process in mitochondria. The systems in endoplasmic reticulum have sometimes been termed "mixed function oxidases" because the terminal enzymes simultaneously oxidize two substrates. In the case of fatty acid desaturation one of these substrates is NADH and the other is the fatty acid. The electrons from NADH are transferred through a specific flavoprotein reductase and a cytochrome to "activate" oxygen so that it will then oxidize the fatty acid. Although the complete mechanism is not worked out, this latter step may involve a hydroxylation. The three components of the system are the desaturase enzyme, cytochrome b_5, and NADH cytochrome b_5 reductase. The overall reaction is

$$R{-}CH_2{-}CH_2{-}(CH_2)_7{-}COOH + NADH + H^+ + O_2 \longrightarrow R{-}CH{=}CH{-}(CH_2)_7{-}COOH + NAD^+ + 2H_2O$$

As noted before, the enzyme specificity is such that the R group must contain at least six carbon atoms. The two main products in most organs are palmitoleic and oleic acids.

The control mechanisms that govern the conversion of the palmitate product of fatty acid synthetase to unsaturated fatty acids are largely unexplored. One of the most important considerations is the control of the proportions of the unsaturated fatty acids available for a proper maintenance of physical state of stored triacylglycerols and membrane phospholipids. A critical committed step in the formation of unsaturated fatty acids from palmitate is the introduction of the first double bond by stearoyl CoA desaturase. The activity of this enzyme and its synthesis are controlled by both dietary and hormonal mechanisms. Increasing the amounts of polyunsaturated fatty acids in the diet of experimental animals decreases the activity of stearoyl CoA desaturase in liver, and insulin, triiodothyronine, and hydrocortisone cause its induction.

Formation of Hydroxy Fatty Acids in Nerve Tissue

There are apparently two different processes that produce α-hydroxy fatty acids in higher animals. One occurs in the mitochondria of many tissues and acts on relatively short-chain fatty acids. This is discussed in Section 9.6. The second process has so far been demonstrated only in tissues of the nervous system where it produces long-chain fatty acids with a hydroxyl group on carbon-2. These are needed for the structure of some myelin lipids. The

specific case of α-hydroxylation of lignoceric acid to cerebronic acid has been studied. These enzymes preferentially use C_{22} and C_{24} fatty acids and show characteristics of the "mixed function oxidase" systems, requiring molecular oxygen and reduced NAD or NADP. This synthesis may be closely coordinated with the biosynthesis of the sphingolipids, which contain the hydroxylated fatty acids.

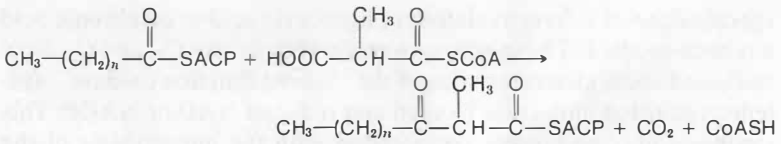
Fatty Acids Formed by Modification of Fatty Acid Synthetase Function

The schemes outlined in previous sections, which utilize palmitate synthesized by fatty acid synthetase and modify it by further enzymatic action, account for the great bulk of fatty acid biosynthesis in the human body, particularly that involved in energy storage. However, there are a number of special instances where smaller amounts of different fatty acids are needed for specific structural or functional purposes, and these acids are produced by modifications of the process carried out by fatty acid synthetase. Two examples are the production of fatty acids shorter than palmitate in mammary glands and the synthesis of branched-chain fatty acids in certain secretory glands.

Recent work has shown that milk produced by many animals contains varying amounts of fatty acids with shorter chain lengths than palmitate. The amounts produced by the mammary gland apparently vary with species and especially with the physiological state of the animal. This is probably true of humans, although most investigations have been carried out with rats, rabbits, and various ruminants. The same fatty acid synthetase that produces palmitate synthesizes the shorter chain acids when the linkage of the growing chain with the acyl carrier protein is split before the full 16-carbon chain is completed. This hydrolysis is caused by soluble thioesterases whose activity is under hormonal control.

As noted in an earlier section, there are relatively few branched-chain fatty acids in higher animals, and, until recently, their metabolism has been studied mostly in primitive species such as *Mycobacteria*, where they are present in greater variety and amount. It is now known that simple branched-chain fatty acids are synthesized by tissues of higher animals for specific purposes, such as the production of waxes in sebaceous glands and avian preen glands and the elaboration of structures in the echo-locating systems of porpoises.

The majority of branched-chain fatty acids in higher animals are simply methylated derivatives of the saturated, straight-chain acids, and they are synthesized by fatty acid synthetase. When methylmalonyl CoA is used as a substrate instead of malonyl CoA, a methyl side chain is inserted in the fatty acid, and the reaction is as follows:



The regular reduction steps then follow. Apparently these reactions occur in many tissues normally at a rate several orders of magnitude slower than the utilization of malonyl CoA to produce palmitate. However, it has been suggested that the proportion of branched-chain fatty acids synthesized is largely governed by the relative availability of the two precursors, and an increase in branching can occur by decreasing the ratio of malonyl CoA to methylmalonyl CoA. A malonyl CoA decarboxylase capable of causing this decrease occurs in many tissues. It has also been suggested that increased levels of methylmalonyl CoA in pathological situations, such as vitamin B_{12} deficiency, can lead to excessive production of branched-chain fatty acids.

Production of Fatty Alcohols

$CH_3—(CH_2)_n—CH_2OH$

Fatty alcohol

As discussed in Chapter 10, many phospholipids contain fatty acid chain moieties in ether linkage rather than ester linkage. The biosynthetic precursors of these ether-linked chains are fatty alcohols rather than fatty acids. These alcohols are formed in higher animals by a two-step, NADPH-linked reduction of fatty acyl CoAs in the endoplasmic reticulum. In organs that produce relatively large amounts of ether-containing lipids, the concurrent production of fatty acids and fatty alcohols is probably closely coordinated.

9.4 STORAGE OF FATTY ACIDS AS TRIACYLGLYCEROLS

Most tissues in the human body can convert fatty acids to triacylglycerols by a common sequence of reactions, but liver and adipose tissue carry out this process to the greatest extent. The latter organ is a specialized connective tissue, which is designed for the synthesis, storage, and hydrolysis of triacylglycerols, and this is the main mechanism that the human body has for relatively long-term energy storage. The triacylglycerols are stored as liquid droplets in the cytoplasm, but this is in no way "dead storage," as turnover

occurs with an average half-life of only a few days. Thus, in a homeostatic situation there is continuous synthesis and breakdown of triacylglycerols in adipose tissue. Some storage also occurs in skeletal and cardiac muscle, but this is only for local consumption.

Triacylglycerol synthesis in the liver is used primarily for production of blood lipoproteins, although the products can serve as energy sources for other liver functions. In all cases the fatty acids for triacylglycerol synthesis come either from the diet or from local biosynthesis. The acetyl coenzyme A for biosynthesis is principally derived from glucose catabolism.

Biosynthesis of Triacylglycerols

Triacylglycerols are synthesized in most tissues from activated fatty acids and a phosphorylated three-carbon product of glucose catabolism (see Figure 9.4). The latter can be either glycerol 3-phosphate or dihydroxyacetone phosphate. Glycerol phosphate is formed either by reduction of dihydroxyacetone phosphate produced in glycolysis or by phosphorylation of glycerol. It is important to note that there is no glycerol kinase in adipose tissue, so in that particular organ glycerol phosphate must be supplied from glycolytic

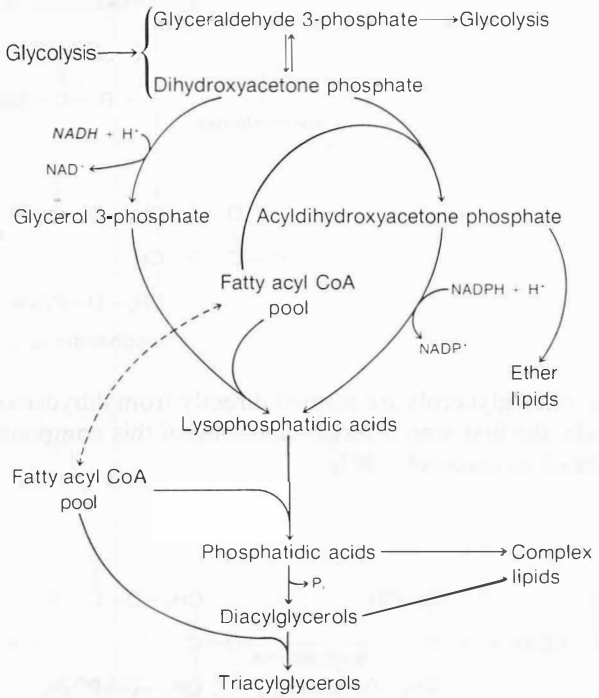
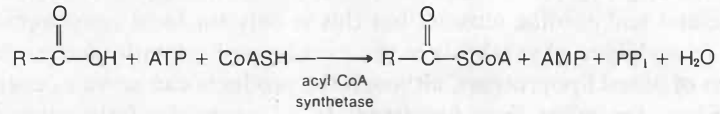


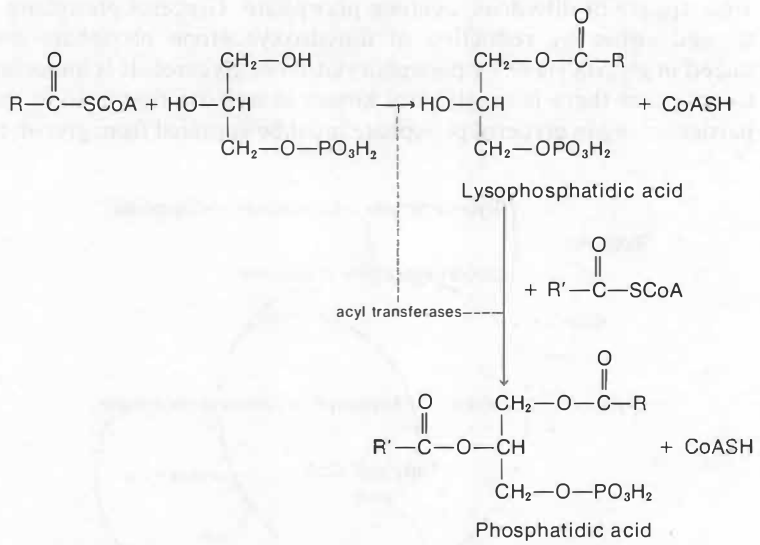
Figure 9.4
Alternative pathways for biosynthesis of triacylglycerols from dihydroxyacetone phosphate.

intermediates. The fatty acids are activated by conversion to their coenzyme A esters in the following reaction:

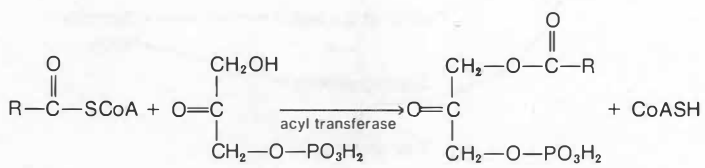


This is a two-step reaction with an acyl adenylate as intermediate and is driven by hydrolysis of the pyrophosphate to inorganic phosphate.

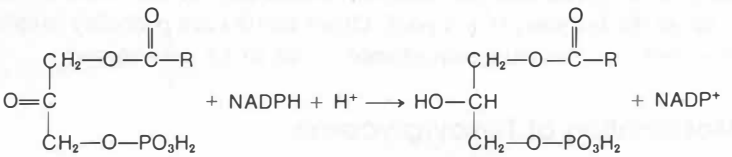
The synthesis of triacylglycerols from the phosphorylated three-carbon fragments involves formation of phosphatidic acid, which is a key intermediate in synthesis of other lipids as well (see Chapter 10). When the esterification occurs with glycerol 3-phosphate, it is simply two sequential acylations.



When triacylglycerols are formed directly from dihydroxyacetone phosphate, the first step is direct acylation of this compound before it is reduced to glycerol 3-PO$_4$.

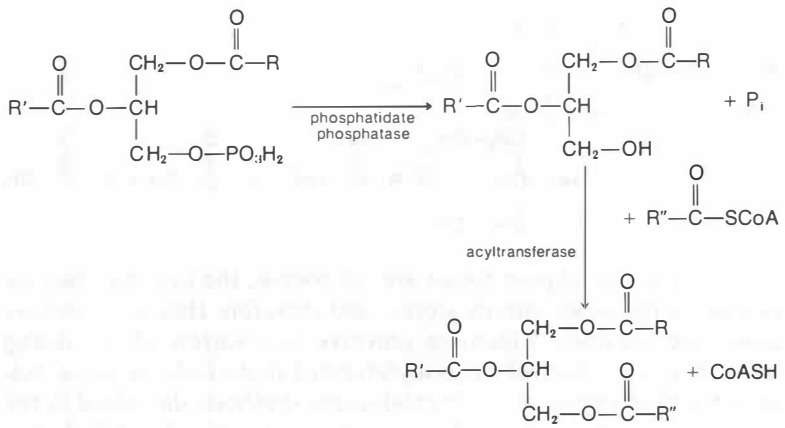


The product, acyldihydroxyacetone phosphate, is then reduced to lysophosphatidic acid, which can be further esterified as shown here. If phosphatidic acid from either of these routes is now to be



Acyldihydroxyacetone phosphate

used for synthesis of triacylglycerols, the phosphate group is next hydrolyzed by phosphatidate phosphatase to yield diacylglycerols. The latter are then acylated to triacylglycerols using a fatty acyl coenzyme A as a source of the fatty acid.



There is at least one tissue, the intestinal mucosa, in which the synthesis of triacylglycerols does not require formation of phosphatidic acid as described above. A major product of intestinal digestion of lipids is 2-monoacylglycerols, which are absorbed as such into the mucosa cells. An enzyme in these cells catalyzes the acylation of these monoacylglycerols with acyl coenzyme A's to form 1,2-diacylglycerols, which then can be further acylated as shown above.

The degree of specificity of the acylation reactions in all the steps above is still quite controversial. Analysis of fatty acid patterns in triacylglycerols from various human tissues shows that the distribution of different acids on the three positions of glycerol is neither random nor absolutely specific. The patterns in different tissues show some characteristic tendencies. Palmitic acid tends to be concentrated in position 1 and oleic acid in positions 2 and 3 of human adipose tissue triacylglycerols. (Position 3 is the one from which

CLIN. CORR. 9.1 OBESITY

When storage of energy as triacylglycerols in adipose tissue becomes abnormal and excessive deposition occurs, the condition is called obesity. In some cases the increased body weight is obvious, but in other patients it is difficult to assess the exact point at which excessive triacylglycerol storage needs to be corrected. Statistics do indicate, however, that obesity is correlated with shorter life span and increased incidence of other health problems.

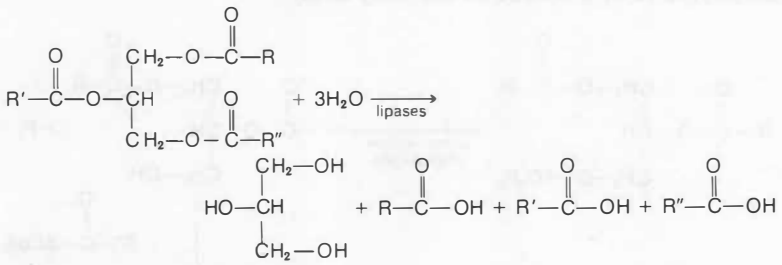
Obesity is a symptom rather than a disease, and many factors lead to it and need to be considered when trying to correct it. The basic cause is, of course, an excess intake of calories as food and drink over and above the energy requirements of the individual for the period involved. Many possible purely biochemical causes for obesity can be hypothesized, such as inadequate lipases to mobilize the triacylglycerols from adipose tissue, but the human body has so many compensating mechanisms to ensure sufficient energy supply to its tissues that such simple causes rarely, if ever, occur. An appetite-controlling center, or centers, in the hypothalamus has been conclusively demonstrated, but the exact details of the peripheral mechanisms that control food intake are yet to be elucidated.

Treatment of obesity is a significant medical problem, although it is frequently treated by nonmedical practitioners. The plethora of available fad and weight-reducing programs shows vividly the economic incentives involved. Responsible treatment for obesity should involve a careful assessment of the individual patient to evaluate the relative importance of

phosphate was removed in hydrolysis of phosphatidic acid.) The two main factors that determine the localization of a given fatty acid to a given position on glycerol are the specificity of the acyltransferase enzyme involved and the relative availability of the different fatty acids in the fatty acyl CoA pool. Other factors are probably involved also, but their relative importance is yet to be determined.

Mobilization of Triacylglycerols

The first step in recovering stored fatty acids for energy production is the hydrolysis of triacylglycerols. A variety of lipases catalyze this reaction, the sequence of hydrolysis from the three positions on glycerol depending upon the specificities of the particular lipases involved.



The lipases in adipose tissue are, of course, the key enzymes for release of the major energy stores, and therefore they are carefully controlled enzymes, which are sensitive to a variety of circulating hormones. This control of triacylglycerol hydrolysis must be balanced with the process of triacylglycerol synthesis described in the previous section to assure adequate energy stores and avoid obesity (see Clin. Corr. 9.1). The fatty acids and glycerol produced by the adipose tissue lipases are released to the circulating blood, where the fatty acids are bound by serum albumin and transported to tissues for use. The glycerol returns to the liver, where it is converted to dihydroxyacetone phosphate and enters the glycolytic pathway.

9.5 METHODS OF INTERORGAN TRANSPORT OF FATTY ACIDS AND THEIR PRIMARY PRODUCTS

The energy available in fatty acids needs to be distributed throughout the body from the site of its absorption, biosynthesis, or storage

to the functioning tissues that consume it. This transport is closely integrated with the transport of other lipids, especially cholesterol. Since these transport systems appear intimately involved in the pathological processes leading to atherosclerosis, they are being intensively studied, but many important questions are still unanswered.

The human body uses three types of substances as vehicles to transport lipid-based energy: (1) chylomicrons and other plasma lipoproteins in which energy in the form of triacylglycerols is carried in protein-coated lipid droplets, the latter also containing other lipids; (2) fatty acids bound to serum albumin; and (3) the so-called "ketone bodies," principally acetoacetate and β-hydroxybutyrate. These three vehicles are used in varying proportions to carry the energy in the bloodstream with three routes. The first is transport of dietary fatty acids throughout the body from the intestine after absorption and uses chylomicrons as a vehicle. The second is the transport of lipid-based energy processed by or synthesized in the liver and distributed either to adipose tissue for storage or to other tissues for utilization. This second process uses "ketone bodies" and plasma lipoproteins other than chylomicrons. The third itinerary is transport of energy released from storage in adipose tissue to the rest of the body, and the vehicle used in this case is fatty acids, which are bound to serum albumin.

the various possible contributing factors. The most logical approach is then to decrease energy input below the needs of the obese individual until excess weight is lost. This decreased dietary intake should be programmed to provide a balance of constituents and a sufficient supply of essential nutrients. A program of increased physical activity within the capabilities of the individual should hasten the process. The use of various weight-reducing drugs is a very controversial alternative.

Forms by Which Lipid-Based Energy Is Transported in Blood

The proportions of energy being transported in any one of the modes outlined above varies considerably with metabolic and physiological state. At any one time, the largest amount of lipid in blood is in the form of triacylglycerols in the various lipoproteins. However, the fatty acids bound to albumin are utilized and replaced very rapidly so the total energy transport for a given period of time by this mode may be very significant.

Blood Lipoproteins

The blood lipoproteins are synthesized in the liver and are a heterogeneous group of lipid–protein complexes composed of various types of lipids and five families of apoproteins. They have been classified empirically according to their size and density as very low density lipoproteins (VLDL), low density lipoproteins (LDL), and high density lipoproteins (HDL) (see Table 9.3), but these distinctions are somewhat arbitrary and overlapping. They are also frequently classified according to electrophoretic mobility into pre-β-, β-, and α-lipoproteins, which correspond generally to the VLDL,

CLIN. CORR. 9.2 GENETIC ABNORMALITIES IN LIPID-ENERGY TRANSPORT

Diseases that affect the transport of lipid-based energy frequently result in abnormally high plasma triacylglycerols, cholesterol, or both. They are classified as hyperlipidemias. Some of them are genetically transmitted, and presumably they result from the alteration or lack of one or more proteins involved in the production or processing of blood lipids. The nature and function of all of these proteins is yet to be determined, so the elucidation of exact causes of the pathology in most of these diseases is still in the early stages. However, in several cases a specific protein abnormality has been associated with altered lipid transport in the patients' plasma.

In the extremely rare disease analbuminemia, there is an almost complete lack of serum albumin. Despite the many functions of this protein, the symptoms of the disease are surprisingly mild. Lack of serum albumin effectively eliminates the transport of fatty acids unless they are esterified in acylglycerols or complex lipids. However, since patients with analbuminemia do have elevated plasma triacylglycerol levels, presumably the deficiency in lipid-based energy transport caused by the absence of albumin to carry fatty acids is filled by increased use of plasma lipoproteins to carry triacylglycerols.

Another genetic defect that has unexpectedly mild consequences is the absence of lipoprotein lipase. The major problem then is the inability to process chylomicrons after a fatty meal. Although pathological fat deposits do occur throughout the

Table 9.3 Characteristics of Plasma Lipoproteins

	Lipoprotein Type			
	Chylomicrons	*Very Low Density Lipoproteins (VLDL)*	*Low Density Lipoproteins (LDL)*	*High Density Lipoproteins (HDL)*
Average % Composition				
Triacylglycerols	86	55	8	5
Phospholipids	8	20	20	30
Cholesterol	2	10	10	5
Cholesterol esters	2	6	37	15
Protein	2	9	25	45
Density[a]	0.92–0.96	0.95–1.006	1.006–1.063	1.063–1.21
Diameter (Å)	10^3–10^4	250–750	200–250	70–120

[a] As compared to water, which has a density set at 1.000. Values are averages; Tables 3.8 and 3.9 contain more extensive values.

LDL, and HDL, respectively. The two most important categories for delivery of lipid-based energy are the chylomicrons and the VLDL, since they contain relatively large amounts of triacylglycerols. The exact precursor–product relationships, if any, between the various types of plasma lipoproteins are yet to be completely defined, as are the roles of the various protein components. It has been proposed that the liver synthesizes VLDL and that the fatty acids from the triacylglycerols are taken up by adipose tissue and other tissues. In the process the VLDL are converted to LDL. The role, if any, of HDL in transport of lipid-based energy is yet to be clarified. All of these lipoproteins are integrally involved in transport of other lipids, especially cholesterol.

The structures of chylomicrons and the various plasma lipoproteins are similar in that they represent liquid droplets of lipid, which are surrounded by polar and/or hydrophilic groups to allow their solubilization or suspension in aqueous medium. The triacylglycerols and cholesterol esters being least polar tend to be in the center of the droplet. The more polar lipids form a shell on the outside of these nonpolar compounds and the proteins associate with the lipid–water interface on the surface. It is probable that the lipid components can interchange between different classes of lipoprotein, and some of the apoproteins probably have functional roles in modifying enzyme ac-

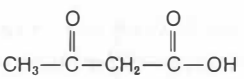
tivity during exchange of lipids between plasma lipoproteins and tissues. Study of some rare genetic abnormalities may help in elucidating the roles of these apoproteins (see Clin. Corr. 9.2).

Fatty Acids Bound to Serum Albumin

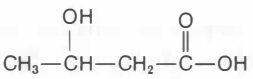
Serum albumin acts as a carrier of a number of substances in the blood, some of the most important being fatty acids. These acids are of course, water insoluble in themselves, but when they are released into the plasma during triacylglycerol hydrolysis they are quickly bound to albumin. This protein has a number of binding sites for fatty acid, two of them with high affinity. At any one time the proportion of sites on albumin actually loaded with fatty acids is far from complete, but the rate of turnover is high, so binding by this mechanism constitutes a major route of energy transfer.

Ketone Bodies

The third mode of transport of lipid-based energy-yielding molecules is in the water-soluble form of small molecules, principally acetoacetate and β-hydroxybutyrate, which are produced from intermediates in the late stages of oxidation of fatty acids. The reactions involved in their formation and utilization will be discussed in a later section



Acetoacetic acid



β-Hydroxybutyric acid

of this chapter. Under certain conditions, these substances can reach excessive concentrations in blood, leading to ketosis and acidosis. When this occurs, some spontaneous decarboxylation of acetoacetate to acetone occurs. This led early investigators to call the group of soluble products "ketone bodies." In fact, these substances are continually produced, primarily by the liver and, to a lesser extent, by the kidney. Skeletal and cardiac muscle then utilize them to produce ATP. Nervous tissue, which obtains almost all of its energy from glucose if it is available, is unable to take up and utilize the fatty acids bound to albumin for energy production. However, it can use β-hydroxybutyrate when glucose supplies are insufficient.

Mechanisms for Transfer of Fatty Acids Between Blood and Tissues

Lipid-based energy distributed as fatty acids bound to albumin or as "ketone bodies" is readily taken up by various tissues for oxidation and production of ATP. However, the energy in fatty acids stored or circulated as triacylglycerols is not directly available, but rather the

body, if patients are put on a low-fat diet they respond reasonably well.

Another rare but more severe disease is caused by loss of one of the apoproteins that form the chylomicrons and VLDL. Under these circumstances the major pathway for transporting lipid-based energy from the diet to the body is unavailable. Chylomicrons, VLDL, and LDL are absent from the plasma and fat absorption is deficient or nonexistent. There are other serious symptoms, including neuropathy and red cell deformities whose etiology is less clear.

Table 9.4 Regulation of Triacylglycerol Metabolism

Enzyme	*Regulatory Agent*	*Effect*
	TRIACYLGLYCEROL MOBILIZATION	
"Hormone-sensitive lipases"	"Lipolytic hormones," e.g., epinephrine, glucagon, ACTH, etc.	Stimulation by cAMP-mediated phosphorylation of relatively inactive enzyme
	Insulin	Inhibition
	Prostaglandins	Inhibition
Lipoprotein lipases	Lipoprotein apoprotein C-II	Activation
	Insulin	Activation
	TRIACYLGLYCEROL BIOSYNTHESIS	
Phosphatidate phosphatase	Steroid hormones	Stimulation by increased enzyme synthesis

latter compounds must be enzymatically hydrolyzed to release the fatty acids and glycerol. There are two types of lipases involved in this hydrolysis: (1) lipoprotein lipases, which hydrolyze triacylglycerols in the blood lipoproteins, and (2) so-called "hormone-sensitive lipases," which initiate the hydrolysis of triacylglycerols in adipose tissue and the release of fatty acids and glycerol into the plasma.

Lipoprotein lipases are located on the surface of the endothelial cells of capillaries and possibly of adjoining tissue cells. They hydrolyze fatty acids from the 1 and/or 3 position of tri- and diacylglycerols when the latter are present in VLDL or chylomicrons. One of the lipoprotein apoproteins must be present to activate the process. The fatty acids that are released are either bound to serum albumin or taken up by the tissue. The monoacylglycerol products may either pass into the cells or be further hydrolyzed by serum monoacylglycerol hydrolase.

A completely distinct type of lipase controls the mobilization of fatty acids from the triacylglycerols stored in the adipose tissue. These lipases are hormonally controlled by a cAMP-mediated mechanism. There are a number of lipase activities in the tissue, but the enzyme attacking triacylglycerols initiates the process. The other lipases then rapidly complete the hydrolysis of mono- and diacylglycerols, releasing fatty acids to the plasma where they are bound to serum albumin (see Table 9.4).

9.6 UTILIZATION OF FATTY ACIDS FOR ENERGY PRODUCTION

The fatty acids that arrive at the surface of tissues are actively taken up by the cells and can be used for energy production. This process occurs primarily inside the mitochondria and is intimately integrated with the processes of energy production from sugars. The energy-rich intermediates produced from fatty acids are the same as those obtained from sugars, that is, reduced pyridine nucleotides, and the final stages of the oxidation process are exactly the same as for carbohydrates, that is, the metabolism of acetyl CoA by the TCA cycle and the production of ATP in the mitochondrial electron transport system.

The degree of utilization of fatty acids for energy production varies considerably from tissue to tissue and depends to a significant degree upon the metabolic status of the body, whether fed or fasted, exercising, and so on. For instance, nerve tissue apparently oxidizes fatty acids to a minimal degree if at all, but cardiac and skeletal muscle depend heavily on fatty acids as a major energy source. During prolonged fasting most tissues are able to use fatty acids for required energy.

β-Oxidation of Straight-Chain Saturated Fatty Acids

For the most part, fatty acids are oxidized by a mechanism that is similar to, but not identical with, a reversal of the process of palmitate synthesis described earlier in this chapter. That is, two-carbon fragments are removed sequentially from the carboxyl end of the acid after steps of dehydrogenation, hydration, and oxidation to form a β-keto acid, which is split by thiolysis. These processes take place while the acid is activated in a thioester linkage to the 4′-phosphopantetheine of coenzyme A.

Activation with Coenzyme A

The first step in oxidation of a fatty acid must therefore be its activation to a fatty acyl CoA. This is the same reaction described for synthesis of triacylglycerols in Section 9.4 and occurs in the endoplasmic reticulum or the outer mitochondrial membrane.

Fatty acids occurring inside the mitochondria can also be activated to a limited extent. This process is analogous to the extramitochondrial one, except that it is dependent on energy from guanine nucleotides instead of adenine nucleotides. The physiological significance of this mitochondrial process is not yet clear.

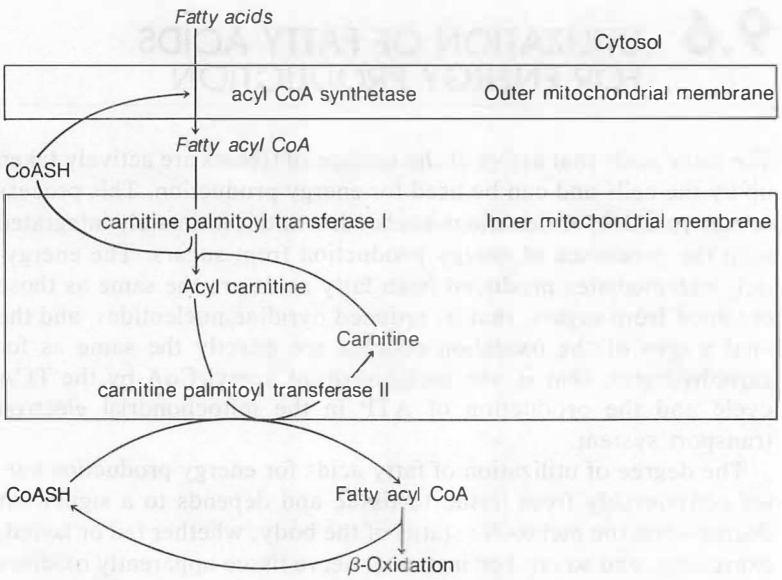
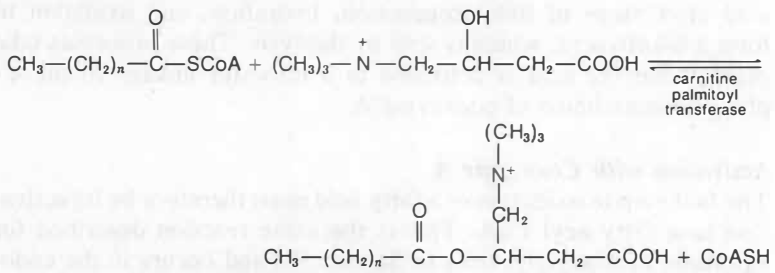


Figure 9.5
Mechanism for transfer of fatty acids from the cytosol through the mitochondrial membranes for oxidation.

Transport of Fatty Acyl CoAs into the Mitochondria

Since most of the fatty acyl CoAs are formed outside the mitochondria while the oxidizing machinery is inside the inner membrane, which is impermeable to coenzyme A and its derivatives, the cell has a major logistical problem. An efficient shuttle system overcomes this problem by using carnitine as the carrier of acyl groups across the membrane. The steps involved are outlined in Figure 9.5. There are enzymes on both sides of the inner mitochondrial membrane, which transfer the fatty acyl group between coenzyme A and carnitine according to the equation:



On the outer surface the acyl group is transferred to carnitine. The acyl carnitine exchanges across the inner mitochondrial membrane with free carnitine. The latter becomes available on the inner surface when the fatty acyl group is transferred back to coenzyme A. Palmitic acid is the major fatty acid transferred in this process and the

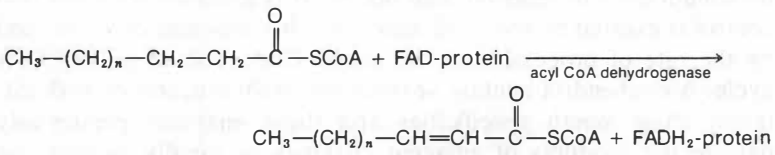
CLIN. CORR. 9.3 GENETIC DEFICIENCIES IN CARNITINE OR CARNITINE PALMITOYL TRANSFERASE

There are several known diseases that are caused by genetic abnormalities in the carnitine system for transporting fatty acids across the inner mitochondrial membrane. They cause deficiencies either in the level of carnitine or in the functioning of the carnitine palmitoyl transferase enzymes. In both cases it is difficult to correlate the degree of the deficiency with the severity of the symptoms.

enzymes have highest affinities for C_{16} acids. Genetic abnormalities in these enzymes lead to muscle pathology (see Clin. Corr. 9.3).

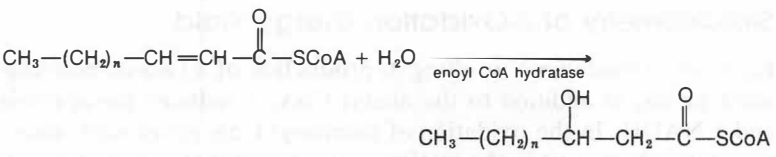
β-Oxidation Reaction Sequence

Once the fatty acyl groups have been transferred back to coenzyme A at the inner surface of the inner mitochondrial membrane they can be oxidized by a group of acyl CoA dehydrogenases in this membrane which remove hydrogens and form a trans double bond between carbons 2 and 3. The several dehydrogenases have different specificities for chain length of the acyl CoA oxidized, and the hydrogen acceptor is a flavoprotein. The reaction is

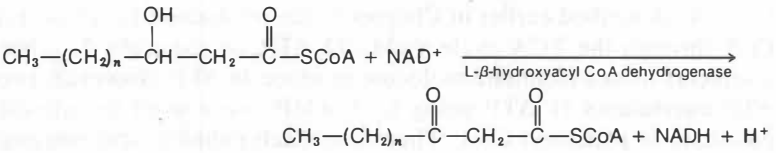


As is the case in the TCA cycle, the enzyme-bound flavoproteins transfer electrons through several other flavoproteins to ubiquinone in the electron transport scheme and only 2 ATP can be obtained for each double bond formed.

The second step in β-oxidation is hydration of the trans double bond to an L-3-hydroxyacyl CoA.



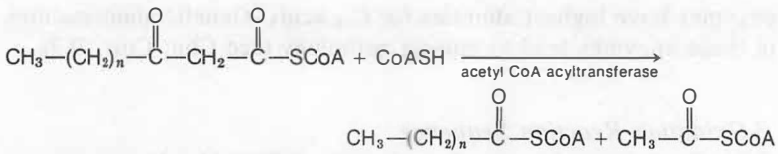
This reaction is stereospecific, in that the L isomer is the product when the trans double bond is hydrated. The stereospecificity of the oxidative pathway is governed by the next enzyme, which is specific for the L isomer as its substrate.



The final step is the cleavage of the two-carbon fragment by a thiolase, which, like the preceding two enzymes, has relatively broad specificity with regard to chain length of the acyl group being oxidized.

The clinical symptoms of carnitine deficiency, in particular, seem to vary greatly. They can range from mild, recurrent muscle cramping to severe ketosis and death. The deficiency has been classified in two categories, that which is a generalized systemic carnitine deficiency, and that which appears to be limited to deficiency only in muscles. Pathological accumulation of triacylglycerols in muscles usually occurs, since fatty acids are inefficiently transported into mitochondria for oxidation. The degree of muscle wasting is variable and results in unpredictable blood carnitine levels, since muscle tissue breakdown causes increased blood carnitine levels.

The deficiency of carnitine palmitoyl transferase is usually a less serious situation and the symptoms are limited to recurrent muscle cramps and aches following strenuous exercise and/or fasting. Muscle mitochondria from patients with this disease have been studied, and the data show that the deficiency can occur with either the carnitine palmitoyl transferase on the inner mitochondrial surface, the one on the outer surface or both. Strangely, triacylglycerols do not accumulate in muscles in this disease.



In the overall process then, an acetyl CoA is produced and the acyl CoA product is ready for the next round of oxidation starting with acyl CoA dehydrogenase.

As yet it has been impossible to show conclusively that any of the enzymes in the β-oxidation scheme are control points, although under rather rigid in vitro conditions some apparently have slower maximum rates of reaction than others. It is generally assumed that control is exerted by the availability of substrates and cofactors and by the rate of processing of the acetyl CoA product by the TCA cycle. Mitochondria contain several acyl dehydrogenases with different chain length specificities and these enzymes presumably oxidize the products of adjacent enzymes as rapidly as they are produced. One way in which substrate availability is controlled is by regulation of the shuttle mechanism that transports fatty acids into the mitochondria. Carnitine palmitoyl transferase which catalyzes the reaction necessary for this transport is inhibited by malonyl CoA, so it can be theorized that the high levels of malonyl CoA present during fatty acid biosynthesis will inhibit β-oxidation.

Stoichiometry of β-Oxidation Energy Yield

Each set of oxidations resulting in production of a two-carbon fragment yields, in addition to the acetyl CoA, 1 reduced flavoprotein and 1 NADH. In the oxidation of palmitoyl CoA seven such cleavages take place, and in the last cleavage 2 acetyl CoAs are formed. The products of β-oxidation of palmitate are thus 8 acetyl CoAs, 7 reduced flavoproteins, and 7 NADH.

Each of the reduced flavoproteins can yield 2 ATP and each of the NADH can yield three when processed through the electron transport chain, so the reduced nucleotides yield 35 ATP per palmitoyl CoA. As described earlier in Chapter 6, the oxidation of each acetyl CoA through the TCA cycle yields 12 ATP, so the eight 2-carbon fragments from a palmitate molecule produce 96 ATP. However, two ATP equivalents (1 ATP going to 1 AMP) were used to activate palmitate to palmitoyl CoA. Therefore, each palmitic acid entering the cell from the action of lipoprotein lipase or from its combination with serum albumin can yield 129 ATP per mole by complete oxidation.

Comparison of the β-Oxidation Scheme with Palmitate Biosynthesis Reactions

In living metabolic systems the reactions in a catabolic pathway are sometimes quite similar to a reversal of the corresponding anabolic pathway, but there are significant differences which provide for separate control of the two schemes. This is true of the palmitate biosynthetic scheme and the scheme for β-oxidation of fatty acids. The critical differences between these two pathways are outlined in Table 9.5. This comparison illustrates some basic mechanisms for separation of metabolic pathways. These include separation by subcellular compartmentation (β-oxidation occurring inside the mitochondria and palmitate biosynthesis in the cytoplasm), and use of slightly different cofactors (NADPH in biosynthesis and NADH in oxidation).

Table 9.5 Comparison of Schemes for Biosynthesis and β-Oxidation of Palmitate

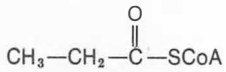
Parameter	*Biosynthesis*	*β-Oxidation*
Subcellular localization	Primarily cytoplasmic	Primarily mitochondrial
Phosphopantetheine-containing active carrier	Acyl carrier protein	Coenzyme A
Nature of small carbon fragment added or removed	Carbons 1 and 2 of malonyl CoA after initial priming	Acetyl CoA
Nature of oxidation–reduction coenzyme	NADPH	FAD when saturated chain dehydrogenated, NAD^+ when hydroxy acid dehydrogenated
Stereochemical configuration of β-hydroxy intermediates	D-β-Hydroxy	L-β-Hydroxy
Energy equivalents yielded or utilized in interconversion of palmitate ↔ acetyl CoA	7ATP + 14NADPH = 49ATP equiv	$7FADH_2$ + 7NADH − 2ATP = 33ATP equiv

Oxidation by Other Processes

The β-oxidation scheme described in the previous sections accounts for the bulk of energy production from fatty acids in the human body. However, it is clear that these reactions must be supplemented by a few other mechanisms so that all fatty acids that are ingested can be oxidized. The principal modifications are those required to oxidize odd-chain fatty acids and unsaturated fatty acids, and those which catalyze α- and ω-oxidation. α-Oxidation occurs at carbon-2 instead of carbon-3 as occurs in the β-oxidation scheme. ω-Oxidation occurs at the methyl end of the fatty acid molecule. Partial oxidation of fatty acids with cyclopropane ring structures probably occurs in humans, but the mechanisms are not worked out.

Odd-Chain Fatty Acids

The oxidation of fatty acids with an odd number of carbons proceeds exactly as described above, but the final product is a molecule of propionyl CoA. In order that this compound can be further oxidized, it undergoes molecular rearrangement and conversion to succinyl CoA. These reactions are identical to those described in Chapter 12 for the metabolism of propionyl CoA when it is formed as a product of the metabolic breakdown of some amino acids.

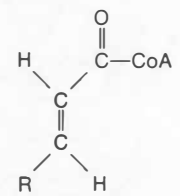


Propionyl CoA

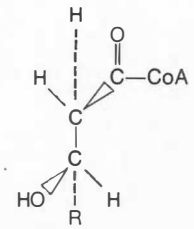
Unsaturated Fatty Acids

The many unsaturated fatty acids in the diet are readily available for the production of energy by the human body. However, in several respects the structures encountered in these dietary acids may differ from those required by the specificity of the enzymes in the β-oxidation pathway. One problem is that the naturally occurring unsaturated fatty acids are almost all of the *cis* configuration, whereas those produced during β-oxidation have the *trans* structure. The enzyme that hydrates double bonds in the β-oxidation process will hydrate *cis* double bonds, but it forms the D-β-hydroxyacyl CoA. In the next step the dehydrogenase that forms the β-ketoacyl structure is stereospecific for the L configuration of the β-hydroxy substrate. However, a racemase enzyme exists in mitochondria, which catalyzes interconversion of D and L isomers. The L isomer is constantly oxidized by the β-hydroxyacyl CoA dehydrogenase, and thus the D isomer is continually converted to the L form and metabolized.

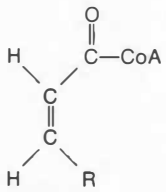
A second problem in β-oxidation of unsaturated fatty acids is that during the process of sequential excision of two-carbon fragments the double bond is sometimes encountered between carbons 3 and 4 instead of between carbons 2 and 3 as required for the β-oxidation scheme enzymes. This problem is overcome by another enzyme,



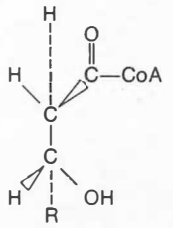
trans-Enoyl fatty acid



L-β-Hydroxy fatty acid



cis-Enoyl fatty acid



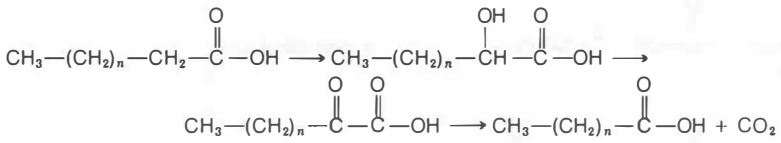
D-β-Hydroxy fatty acid

which converts the *cis* bond between carbons 3 and 4 to a *trans* bond between carbons 2 and 3. The regular scheme can then proceed.

Both of these problems are encountered in the oxidation of linoleoyl coenzyme A, which is outlined in Figure 9.6.

α-Oxidation

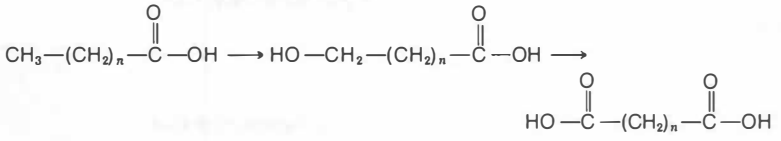
As noted in the earlier discussion of fatty acid biosynthesis, there are several mechanisms for hydroxylation of fatty acids. The one discussed previously is for α-hydroxylation of the long-chain acids needed for the structures of sphingolipids. In addition, there are systems in other tissues which hydroxylate the α-carbon of shorter chain acids in order to start their oxidation. The sequence is as follows:



These hydroxylations probably occur in the endoplasmic reticulum and mitochondria and involve the "mixed function oxidase" type of mechanism discussed previously, because they require molecular oxygen, reduced pyridine nucleotides and specific cytochromes. These reactions are particularly important in oxidation of methylated fatty acids (see Clin. Corr. 9.4).

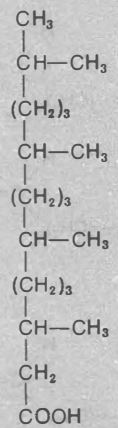
ω-Oxidation

Another minor pathway for fatty acid oxidation also involves hydroxylation and occurs in the endoplasmic reticulum of many tissues. In this case the hydroxylation takes place on the methyl carbon at the other end of the molecule from the carboxyl group or on the carbon next to the methyl end. It also uses the "mixed function oxidase" type of reaction requiring cytochrome P_{450}, O_2, and NADPH as well as the necessary enzymes. The hydroxylated fatty acid can be further oxidized to a dicarboxylic acid, and the β-oxidation can proceed from either end of the molecule. This process probably occurs primarily with medium-chain length fatty acids, and the degree to which it contributes to fatty acid oxidation depends on the tissue and its metabolic state. The overall reactions are



CLIN. CORR. 9.4
REFSUM'S DISEASE

Although the use of the α-oxidation scheme is a relatively minor one in terms of total energy production, it is significant in the metabolism of dietary fatty acids which are methylated. A principal one of these is phytanic acid,



Phytanic acid

a metabolic product of phytol, which occurs as a constituent of chlorophyll. Phytanic acid is a significant constituent of milk lipids and animal fats, and normally it is metabolized by an initial α-hydroxylation followed by dehydrogenation and decarboxylation. β-Oxidation cannot occur initially because of the presence of the 3-methyl group, but it can proceed after the decarboxylation. The whole reaction produces three molecules of propionyl CoA, three molecules of acetyl CoA, and one of isobutyryl CoA.

In a rare genetic disease called Refsum's disease, the patients lack the α-hydroxylating enzyme and accumulate large quantities of phytanic acid in their tissues and serum. This leads to serious neurological problems for reasons which are still obscure.

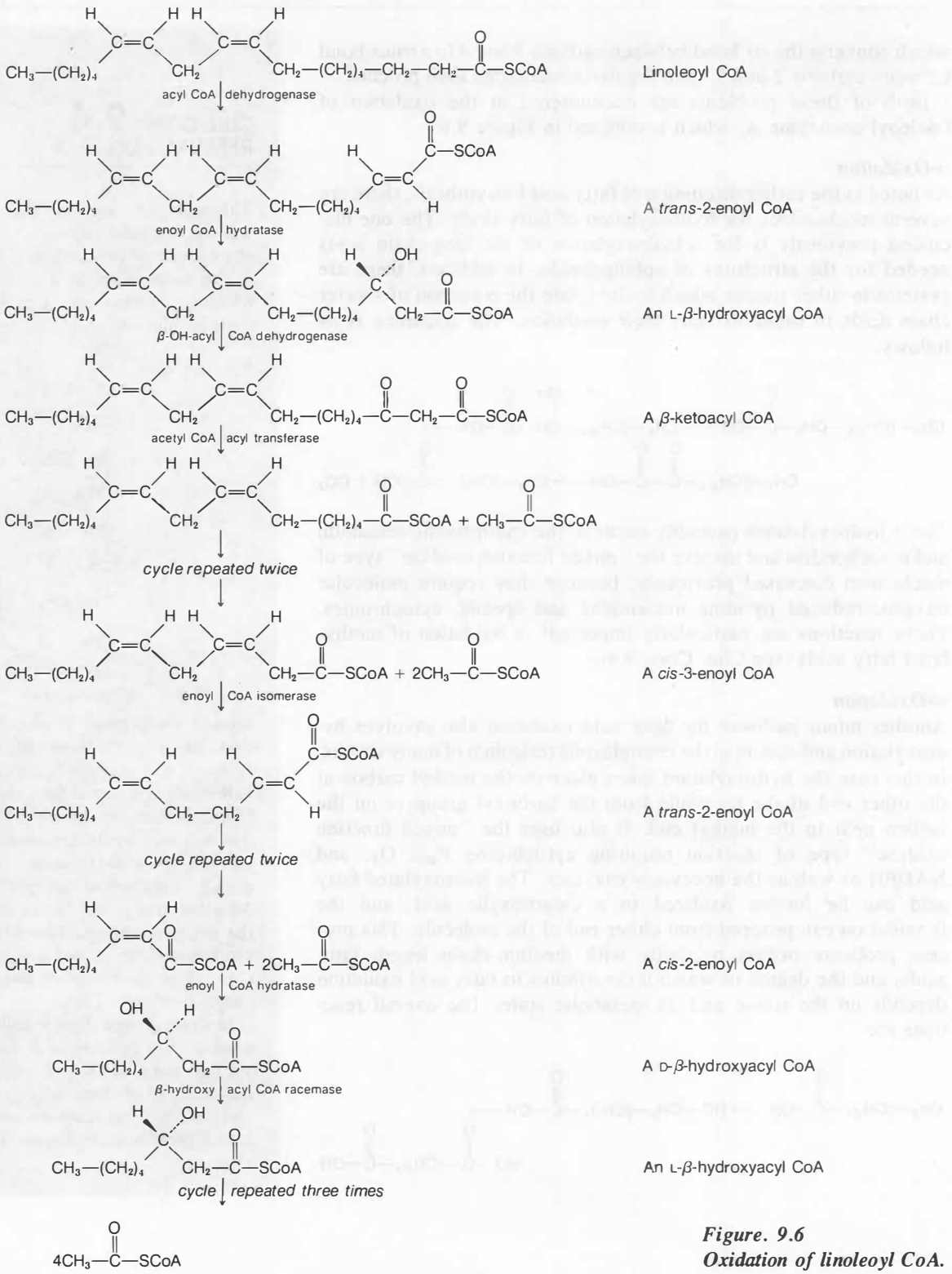


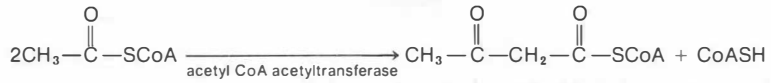
Figure. 9.6
Oxidation of linoleoyl CoA.

Ketone Body Formation and Utilization

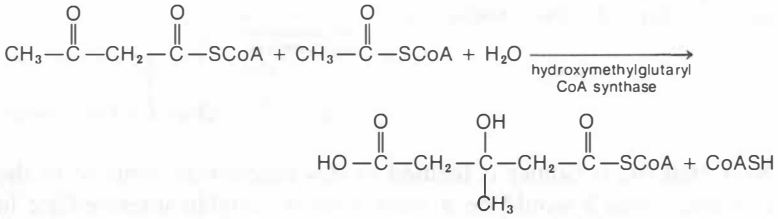
As noted previously, the so-called ketone bodies, which are the most water-soluble form of lipid-based energy, consist mainly of acetoacetic acid and β-hydroxybutyric acid. The latter is a reduction product of the former. Acetoacetyl CoA is an intermediate near the end of the β-oxidation sequence, and it was initially presumed that enzymatic removal of coenzyme A at this point was the main route for production of acetoacetic acid. However, more definitive data subsequently indicated that β-oxidation proceeds completely to acetyl CoA production without accumulation of any intermediates, and acetoacetate is formed afterward from acetyl CoA by a separate mechanism. This process is also integrated with the initial steps in the production of cholesterol.

Formation of Acetoacetate

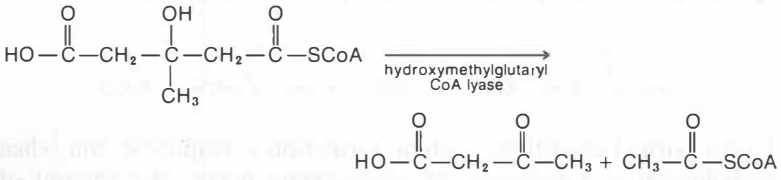
The primary site for the formation of ketone bodies is liver and lesser activity occurs in kidney. The first reaction is condensation of two acetyl CoA molecules to form acetoacetyl CoA.



Acetoacetyl CoA then condenses with another molecule of acetyl CoA to form β-hydroxy-β-methylglutaryl coenzyme A (HMG CoA), which is an important precursor for cholesterol biosynthesis (see Figure 9.7).



Cleavage of HMG CoA then yields acetoacetic acid and acetyl CoA.



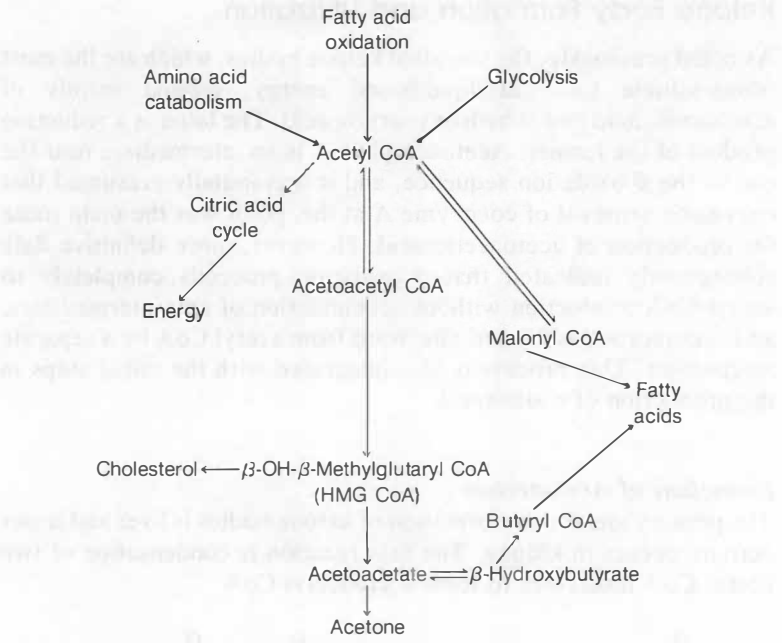
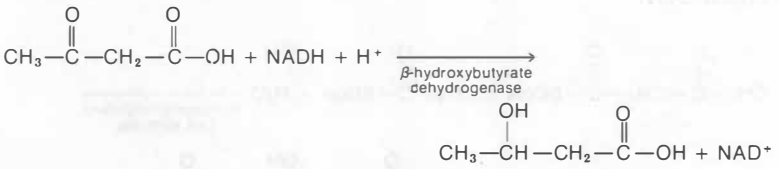


Figure 9.7
Interrelationships of ketone bodies with lipid, carbohydrate, and amino acid metabolism.

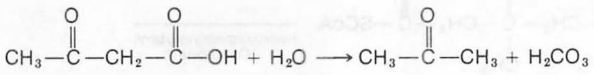
Formation of β-Hydroxybutyrate and Acetone

In mitochondria a proportion of acetoacetate is reduced to β-hydroxybutyrate depending on the intramitochondrial NAD/NADH ratio.



Note that the D isomer is formed in this reaction in contrast to the L isomer, which would be a coenzyme A-coupled intermediate in β-oxidation.

A certain amount of acetoacetate is continually undergoing slow, spontaneous nonenzymatic decarboxylation to acetone.



Under normal conditions acetone formation is negligible, but when pathological accumulations of acetoacetate occur, the amount of

acetone in blood can be sufficient to cause it to be detectable in a patient's breath.

Utilization of Ketone Bodies

Production of acetoacetate and β-hydroxybutyrate is a continuous normal process in the human body. Several tissues, especially cardiac and skeletal muscle, actively absorb these substances and use them as energy sources. At least some parts of the nervous system also utilize these substances when blood glucose levels decrease, but liver is apparently unable to do so. The mechanism for utilization of ketone bodies first requires reactivation of acetoacetate. This is accomplished by a mitochondrial enzyme that uses succinyl CoA as the source of the coenzyme. Acetoacetyl CoA is then used for energy as it would be when formed as a product of β-oxidation. Mitochondrial β-hydroxybutyrate dehydrogenase also reconverts β-hydroxybutyrate to acetoacetate as the concentration of the latter is decreased by oxidation.

$$\text{Acetoacetate} + \text{succinyl CoA} \longrightarrow \text{acetoacetyl CoA} + \text{succinate}$$

In addition to supplying energy by these reactions ketone bodies can also act as precursors for lipogenesis in some tissues. β-Hydroxybutyrate is converted to butyryl CoA and used as a primer for fatty acid synthesis by the reactions shown in Section 9.3. Cytoplasmic enzymes also can convert acetoacetate back to acetoacetyl CoA and acetyl CoA. The latter serve as substrates for biosynthesis of various lipids.

Ketoacidosis

A major clinical problem occurs when the processes of homeostasis are disrupted to the extent that excessive accumulations of ketone bodies occur. This can be caused by a number of pathological situations, one of the most common being uncontrolled diabetes mellitus. When abnormally high levels of ketone bodies occur (ketosis), the level of reabsorption by the kidney is exceeded and ketonuria results. Abnormal accumulation also leads to metabolic acidosis (see Chapter 23), since the two major ketone bodies are organic acids with pK's lower than carbonic acid, the major physiological buffer. This state is then called ketoacidosis. If uncontrolled, this condition leads to rapid deterioration and death.

The factors that lead to ketoacidosis are very complex and are related to the balance between the various mechanisms that provide energy to the body. This balance is normally maintained by hormonal control mechanisms as well as by the relative availability of

the several major energy substrates, glucose, fatty acids, and amino acids. When glucose availability is markedly decreased one physiological response is to increase the mobilization of triacylglycerols stored in adipose tissue. There is a corresponding increase in the oxidation of fatty acids by the tissues for energy production. Synthesis of acetoacetate and β-hydroxybutyrate by the liver is also stimulated and these substances are utilized by peripheral tissues to an increased degree.

In the physiological situation of prolonged fasting, these shifts in energy production from one substrate to another take place smoothly so that moderate physiological ketosis occurs without induction of acidosis. On the other hand, in diabetes mellitus a lack of appropriate levels of functioning insulin disrupts these homeostatic shifts. Insulin is required for uptake of glucose from the blood by tissues, so despite the presence of glucose in the blood the body markedly increases the mobilization of triacylglycerols from adipose tissue to supply the tissues with a substrate for energy production. Plasma levels of fatty acids bound to albumin rise greatly. These high levels of fatty acids, together with a marked acceleration in the rate of fatty acid oxidation within the liver, result in abnormally high acetyl CoA formation in that organ. The normal pathways for utilization of acetyl CoA by liver, that is, oxidation by the TCA cycle and synthesis of fatty acids, become saturated, and excess liver acetyl CoA is shifted to the production of ketone bodies. Since the hormonal signals throughout the body (low insulin and high glucagon; see Chapter 14) are acting to decrease the storage of the excess fatty acids as adipose tissue triacylglycerols, the condition is further aggravated. The amount of ketone bodies produced by liver then becomes sufficient to overcome the ability of the tissues to utilize them. Their buildup reaches levels above the ability of physiological buffers to compensate and acidosis occurs.

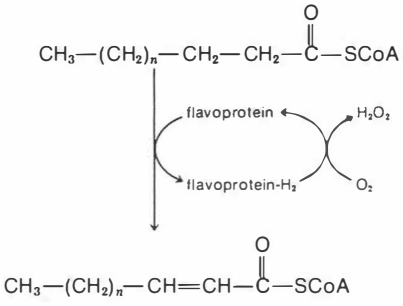
A complete discussion of the hormonal and other factors involved in ketoacidosis is inappropriate in this context. Suffice it to say that a complete understanding of the roles and interactions of these factors is yet to be achieved. It should be noted, however, that ketoacidosis can be encountered in a variety of clinical conditions. The main cause is an overproduction or an underutilization of acetoacetate and/or β-hydroxybutyrate, or a combination of these. If untreated it is a terminal condition.

Putative Role of Peroxisomes in Fatty Acid Oxidation

Most of the oxidation of fatty acid probably occurs in mitochondria, but recent experimental evidence has led to the hypothesis that sig-

nificant oxidation of fatty acids also takes place in the peroxisomes of liver, kidney, and other tissues. Peroxisomes are a class of subcellular organelles with distinctive morphological and chemical characteristics. Their initial distinguishing characteristic was a high content of the enzyme catalase, and it has been suggested that peroxisomes may function in a protective role against oxygen toxicity. Two lines of evidence suggest that they may also be involved in lipid catabolism. First, the analogous structures in plants, glyoxysomes, are capable of oxidizing fatty acids, and, second, a number of drugs used clinically to decrease triacylglycerol levels in patients cause a marked increase in histologically detectable peroxisomes. Subsequently, it has been shown conclusively that liver peroxisomes, isolated by differential centrifugation, do oxidize fatty acids, and do possess most of the enzymes in the β-oxidation scheme.

The mammalian peroxisomal fatty acid oxidation scheme is similar to that in plant glyoxysomes, and it differs from the mitochondrial β-oxidation scheme only in that the initial dehydrogenation is accomplished by a mixed-function oxidase system and is cyanide-insensitive. This step can be described as shown. The hydrogen peroxide is then eliminated by catalase, and the remaining steps are the same as in the mitochondrial system. There is evidence that the peroxisomal and mitochondrial enzymes are slightly different and that the specificity in peroxisomes is for somewhat longer chain length. For this reason the peroxisomes can be viewed as functioning to shorten the chains of relatively long fatty acids prior to mitochondrial oxidation. The extent to which peroxisomal oxidation contributes to the total fatty acid oxidation is not yet known.



9.7 CHARACTERISTICS, METABOLISM, AND FUNCTIONAL ROLE OF POLYUNSATURATED FATTY ACIDS

In recent years there has been considerable renewed interest in elucidating the specific physiological roles of the polyunsaturated fatty acids at the biochemical level. This is due to some extent to the results of initial studies which suggested that a diet in which the proportion of polyunsaturated to saturated fatty acids was relatively high could help to lower blood cholesterol levels in some patients. The relationship between these diet modifications and the develop-

ment of atherosclerosis, if any, is not simple, but the initial reports did tend to spur interest in the polyunsaturated fatty acids.

It is important to emphasize that the essential fatty acids mentioned in Section 9.3 are polyunsaturated but that most of the individual polyunsaturated fatty acids need not be specifically supplied in the diet. All fatty acids with three or more double bonds, without regard to their position in the chain, are polyunsaturated and most of these can be synthesized by human tissues. The essential acids refer to only the linoleic and the linolenic series, which have double bonds near the methyl end of the chain, and which cannot be synthesized by humans. The degree to which these acids really are essential for humans is still to be determined, although clearly reproducible deficiency states can be produced in rats and some other animals by carefully controlled diets. The need for the linoleic acid series is clearer since the discovery of the prostaglandins, which are derived from arachidonic acid, one of the linoleic series. The need for the linolenic acid series is very obscure, although the 22-carbon hexaenoic acid derivative of it shown in Section 9.2 is concentrated in some membranes of nerve and retina. A deficiency syndrome for linolenate has yet to be produced in any animal except rainbow trout, but this is possibly due to extremely efficient mechanisms for conservation of linolenate and its derivatives in the body.

A major role of all polyunsaturated fatty acids seems to be to produce the proper fluidity in biological membranes. As described in Chapter 10, the various phospholipids have variable amounts of polyunsaturated fatty acids as constituents, and it has been conclusively demonstrated that lower organisms can alter the fatty acid patterns in their membrane phospholipids to maintain proper fluidity under changing conditions, such as temperature alterations. This can be done by increasing the proportion of fatty acids with a few double bonds in them or by increasing the degree of unsaturation of the fatty acids.

Metabolic Modifications of Unsaturated Fatty Acids

The human body can synthesize a variety of polyunsaturated fatty acids by the elongation and desaturation reactions described in Section 9.3. The stearoyl CoA desaturase introduces an initial double bond between carbons 9 and 10 in a saturated fatty acid, and then double bonds can be introduced just beyond carbons 4, 5, or 6. Desaturation at carbon-8 probably occurs also in some tissues. The positions of these desaturations are shown in Figure 9.8. The relative specificities of the various enzymes are still to be elucidated completely, but it seems likely that elongation and desaturation can

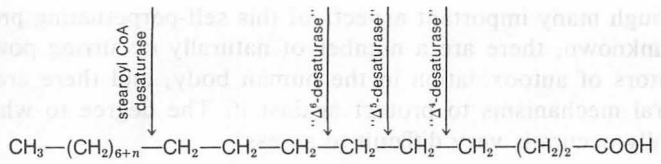
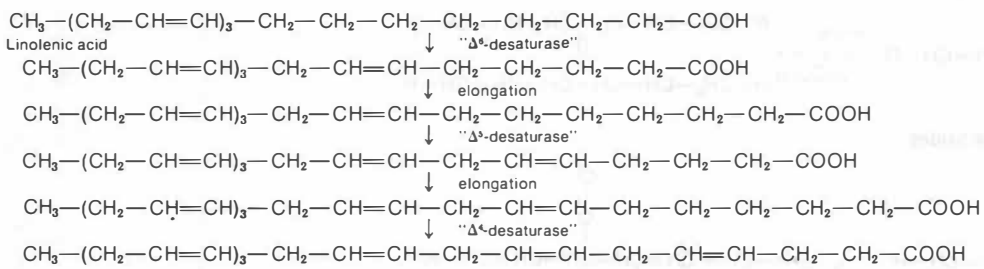


Figure 9.8
Positions in the fatty acid chain where desaturation can occur in the human body.
There must always be at least six single bonds in the chain toward the methyl end of the molecule just beyond the bond being desaturated.

occur in either order. The conversion of linolenic acid to all *cis*-4, 7, 10, 13, 16, 19-docosahexaenoic acid in brain is a specific example of such a sequence.



All-*cis*-4,7,10,13,16,19-docosahexaenoic acid

Formation of Prostaglandins

The polyunsaturated fatty acids, particularly arachidonic acid, are the precursors of the highly active prostaglandins and thromboxanes. A number of different classes of prostaglandins are formed depending on the precursor fatty acid and the sequence of various oxidations which convert the acids to the active compounds. A detailed discussion of these substances and their formation is given in Chapter 10.

Autooxidation of Polyunsaturated Fatty Acids

Polyunsaturated fatty acids in living systems have a significant potential for autooxidation, a process that may have important physiological and/or pathological consequences. This is the set of reactions that causes rancidity in fats and the curing of linseed oil in paints.

Although many important aspects of this self-perpetuating process are unknown, there are a number of naturally occurring potential initiators of autooxidation in the human body, and there are also several mechanisms to protect against it. The degree to which it actually occurs is very difficult to assess.

Autooxidation Reactions

The reactions probably involved in lipid autooxidation are shown in Figure 9.9. The basis behind the process is that the methylene carbon between any two double bonds in the polyunsaturated fatty acids is quite susceptible to hydrogen abstraction and free radical formation. Once this abstraction occurs the reactions can take place

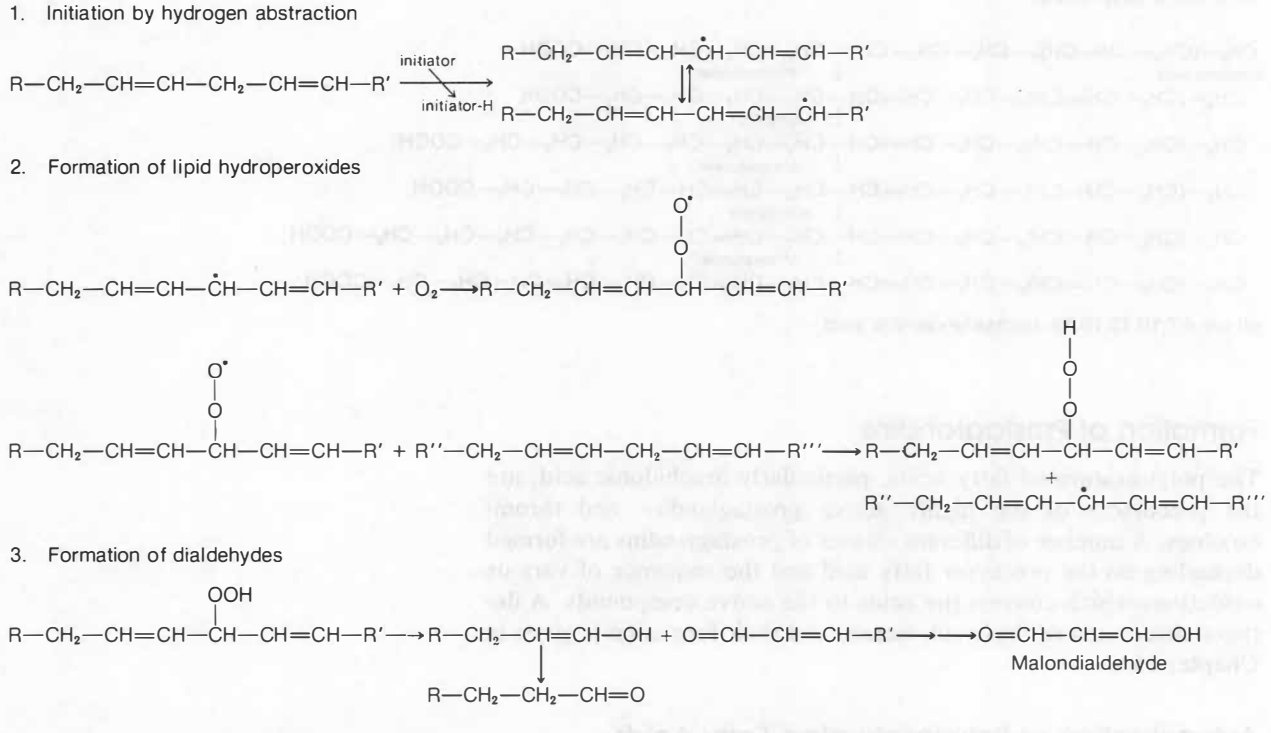


Figure 9.9
Possible reactions and typical products of fatty acid autooxidation.

in any sequence, and many of the reactive breakdown products can contribute to further oxidation. Oxygen can attach to acids from which hydrogen has been abstracted, forming free radicals which can then react with another lipid molecule, leading to abstraction of hydrogen from the second molecule. The products of this reaction are a lipid hydroperoxide in the first molecule and a new free radical in the molecule attacked. The lipid hydroperoxide molecules break up, forming dialdehydes, the most prominent being malondialdehyde. This product can cause cross-linking between various types of molecules, such linkages leading to cytotoxicity, mutagenicity, membrane breakdown, and enzyme modification. Malondialdehyde also polymerizes with itself and other tissue breakdown products, forming an insoluble pigment, lipofuscin, which accumulates in some aging tissues.

Possible Autooxidation Initiators

A number of external agents can initiate autooxidation in vitro. The extent to which they can lead to such reactions in vivo in humans is undetermined. Various types of radiation, including sunlight, and environmental pollutants such as oxides of nitrogen and carbon tetrachloride are examples of such external agents. The detoxification mechanisms for CCl_4 in the liver use cytochrome P_{450} and generate transient free radicals. The latter can initiate lipid autooxidation and lead to carcinogenesis. Metabolism of the herbicide paraquat, sometimes used for marijuana control, produces superoxide anions, which can also initiate fatty acid autooxidation.

It is quite possible that autooxidation can be initiated without the need for an external agent. Theoretically at least, the enzymes involved in various oxidative processes can produce singlet oxygen and transient partial reduction products of oxygen (superoxide anion, hydrogen peroxide, and the hydroxyl radical), any or all of which could potentially lead to lipid free radicals and/or lipid hydroperoxides. For instance, under proper circumstances rat liver microsomes cause extensive formation of lipid peroxides in vitro, presumably initiated by an enzyme-bound reactive form of iron. A number of enzymes such as xanthine oxidase, superoxide dismutase, and lipoxidases can initiate lipid peroxide formation in vitro.

Possible Protective Mechanisms in Vivo

Recent evidence suggests that in special circumstances the partial reduction products of oxygen which can potentially initiate lipid autooxidation may actually be produced for beneficial purposes, for example, by leukocytes in killing bacteria; however, under most conditions the human body utilizes potent mechanisms to ensure

against accumulation of these substances. Three types of enzymes, the catalases, peroxidases, and superoxide dismutases, seem primarily designed to destroy them rapidly and keep tissue levels of their substrates negligible. An additional safeguard is the presence of scavenging molecules in the body, which interact with any free radicals produced, but do not in turn produce self-propagating chain reactions. The principle one present in humans is α-tocopherol, vitamin E. Evidence for its protective role in vivo is purely circumstantial.

Potential Pathological Consequences of Fatty Acid Autooxidation

The two pathological processes that are most likely to be associated with the oxidative breakdown of polyunsaturated fatty acids are the development of cancer and the degradation of tissues during aging. It has long been assumed that free radicals can act as carcinogens by altering DNA molecules and causing mutations. The free radicals formed during lipid autooxidation could thus be carcinogenic, although direct evidence for their involvement is lacking. DNA can also be altered by other types of compounds besides free radicals, and some products of fatty acid autooxidation, for instance malondialdehyde, have been shown to be cytotoxic and mutagenic to some tissues in culture.

It has been suggested that continued long-term accumulation of tissue injury from free radical reactions eventually leads to functional deterioration of tissues and to the symptoms associated with the aging process. The possibility that the autooxidation of fatty acids may be involved in this deterioration is strengthened by the well-established accumulation with age of the pigment lipofuscin in human tissues such as brain and testis. This pigment is probably an unmetabolizable product of the lipid autooxidation process. However, no direct pathology has been shown to result from its accumulation.

There is further preliminary evidence that may relate fatty acid autooxidation to aging from studies of the possible effects of dietary vitamin E on the life span of mice and from the analysis of polyunsaturated fatty acid patterns in brain membranes at different ages. Further work is needed before definite conclusions can be drawn from this work.

BIBLIOGRAPHY

Dietschy, J. M., Gotto, A. M., Jr., and Ontko, J. A. (eds.). *Disturbances in lipid and lipoprotein metabolism.* Bethesda, Md.: American Physiological Society, 1978.

Gurr, M. I., and James, A. T. *Lipid biochemistry, an introduction,* 3rd ed. London: Chapman and Hall, 1980.

IUPAC-IUB Commission on Biochemical Nomenclature. The nomenclature of lipids. *Biochem. J.* 171:21, 1978.

Lech, J. J. (chmn.). Symposium on control of endogenous triglyceride metabolism in adipose tissue and muscle. *Fed. Proc.* 36:1984, 1977.

McGarry, J. D., and Foster, D. W., Regulation of hepatic fatty acid oxidation and ketone body production. *Annu. Rev. Biochem.* 49:395, 1980.

Nilsson-Ehle, P., Garfinkel, A. S., and Schotz, M. C. Lipolytic enzymes and plasma lipoprotein metabolism. *Annu. Rev. Biochem.* 49:667, 1980.

Robinson, A. M., and Williamson, D. H., Physiological roles of ketone bodies as substrates and signals in mammalian tissues. *Physiol. Rev.* 60:143, 1980.

Scow, R. O., Blanchette-Mackie, E. J., and Smith, L. C. Transport of lipid across capillary endothelium. *Fed. Proc.* 39:2610, 1980.

Volpe, J. J. Lipid metabolism: fatty acid and cholesterol biosynthesis. *Adv. Mod. Nutr.* 37, 1978.

10

Lipid Metabolism II

ROBERT H. GLEW

PATHWAYS OF METABOLISM OF SPECIAL LIPIDS

10.1 OVERVIEW

Lipid is a general term that describes substances that are relatively water-insoluble and extractable by nonpolar solvents such as chloroform, benzene, petroleum ether, and diethyl ether. The complex lipids of man fall into one of two broad categories: the nonpolar and the polar. Examples of nonpolar lipids are triacylglycerols and cholesterol esters. The polar lipids are amphipathic in that they contain both a hydrophobic domain and a hydrophilic region in the same molecule. The two major subdivisions of the polar lipids are the phospholipids and the sphingolipids. The hydrophobic and hydrophilic domains are bridged by a glycerol moiety in the case of glycerophospholipids and by sphingosine in the case of sphingomyelin and the glycosphingolipids. In terms of location, triglyceride is confined largely to storage sites in adipose tissue, whereas the polar lipids occur primarily in cellular membranes. Membranes generally contain about 40% of their dry weight as lipid and 60% as protein, the two being held together by noncovalently bonded interactions.

The fundamental function of membranes is to provide compartments that separate two environments that have different solute compositions and concentrations. Specialized transfer systems usually involving proteins allow for selective and controlled passage of substances across the membrane from one compartment to another.

Membranes perform a second important function: that of providing structural supports for complex biochemical processes involving a long series of coordinated reactions. The various enzymes involved in a sequence of related reactions are often bound together in an orderly fashion to the surface of a membrane. The membrane maintains a constant spatial relationship between the enzymes such that the product of one reaction is made available as the substrate to the second enzyme in the pathway—a condition that cannot be achieved when the enzymes and substrates are free in solution. Polar lipids play an important role in the association of proteins and enzymes with membranes.

A third major function of membranes—the plasma membrane specifically—is to permit cells to recognize each other. The processes of phagocytosis, contact inhibition, and rejection of transplanted tissues and organs are all phenomena of medical significance that involve highly specific recognition sites on the surface of the plasma membrane. Complex glycosphingolipids appear to play a role in these important biological events.

The polar lipids, including both phospholipids and sphingolipids, are amphipathic in character, and self-association of lipid molecules

occurs by means of intermolecular hydrophobic interactions. Nearly all models of membrane structure incorporate a lipid bilayer arrangement where the long aliphatic chains are sequestered within the interior of the membrane structure and the polar, hydrophilic head groups face the aqueous environment. Because hydrophobic groups in water tend to destabilize the structure of water, the apolar groups cluster together, thereby excluding water and reducing the free energy of the system. The detailed models of membrane structure are discussed elsewhere (Chapter 5). As we shall see, the hydrophobic tails of the polar lipids are made up of the long aliphatic chains of fatty acids or sphingosine derivatives.

The glycolipids are worthy of study because the ABO antigenic determinants of the blood groups are primarily glycolipid in nature. In addition, various sphingolipids are the storage substances that accumulate in the liver, spleen, kidney, or nervous tissue of persons suffering from certain genetic disorders called sphingolipidoses. In order to understand the basis of these enzyme-deficiency states, a knowledge of the relevant chemical structures involved is required.

10.2 THE ACYLGLYCEROLIPIDS–TRIACYLGLYCEROLS AND PHOSPHOLIPIDS

The two principal classes of acylglycerolipids are the triacylglycerols and glycerophospholipids. They are referred to as glycerolipids because the core of these compounds is provided by the 3-carbon polyol, glycerol.

The two primary alcohol groups of glycerol are not stereochemically identical and in the case of the phospholipids, it is usually the same hydroxyl group that is esterified to the phosphate residue. The stereospecific numbering system is the best way to designate the different hydroxyl groups. In this system, when the structure of glycerol is drawn in the Fischer projection with the C-2 hydroxyl group projecting to the left of the page, the carbon atoms are numbered as shown in Figure 10.1. When the stereospecific numbering (sn) system is employed, the prefix "sn-" is used before the name of the compound. Glycerophospholipids usually contain a sn-glycerol 3-phosphate moiety. Although each contains the glycerol moiety as a fundamental structural element, the neutral triacylglycerols and the charged, ionic phospholipids have very different physical properties and functions. Let us now consider the structure and function of the various glycerolipids.

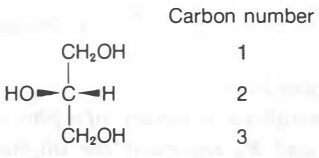


Figure 10.1
Stereospecific numbering of glycerol.

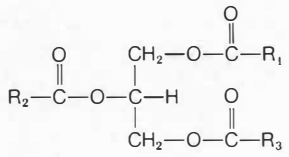


Figure 10.2
Generalized structure of a triacylglycerol molecule where $-\overset{O}{\overset{\|}{C}}-R$ *represents a fatty acid residue.*

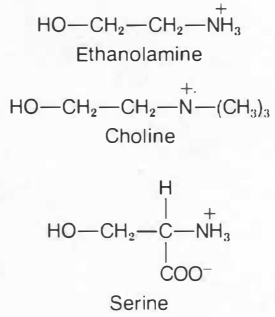


Figure 10.4
The structures of some common polar head groups of phospholipids.

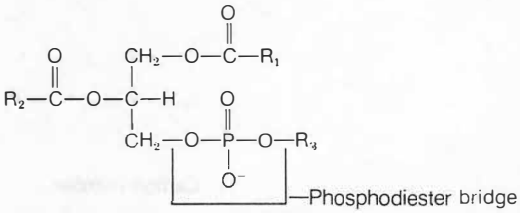


Figure 10.5
Generalized structure of a phospholipid where R_1 *and* R_2 *represent the aliphatic chains of fatty acids, and* R_3 *represents some polar head group.*

Structure of Acylglycerolipids

Triacylglycerols

The 1,2,3-triacylglycerols represent a storage form of fatty acids in which each of the three alcohol groups of glycerol are esterified to fatty acids (Figure 10.2).

Triacylglycerols, or triglycerides as they are sometimes called, are referred to as fats or oils depending on whether they are solid or liquid, respectively, at room temperature. The physical properties of a particular triacylglycerol are determined by the nature of the R groups, or fatty acids attached to the glycerol backbone. The structures of some of the common fatty acids that comprise the triacylglycerols and phospholipids are indicated on page 443. The longer the chain length and the fewer the number of carbon–carbon

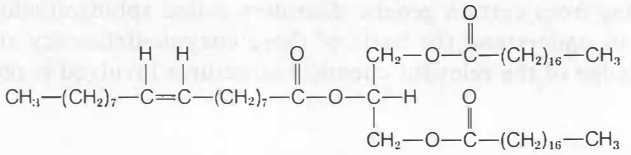


Figure 10.3
A triglyceride containing stearic acid (18 : 0) in the sn-1 and sn-3 positions and oleic acid [18 : 1 (Δ9)] in the sn-2 position.

double bonds in the fatty acid moieties, the higher will be the melting point of the triacylglycerol that contains these fatty acids. In contrast, triacylglycerols composed of relatively short-chain length fatty acids (e.g., butyric acid) and polyunsaturated fatty acids (e.g., arachidonic and linolenic acids) have lower melting points and usually occur as liquids at room temperature. The longer chain length, more saturated or highly reduced fatty acids are located in the R_1 and R_3 positions, whereas shorter, more polar, unsaturated fatty acids occupy the internal R_2 position (Figure 10.3).

Phospholipids

The phospholipids are polar, ionic lipids composed of 1,2-diacylglycerol and a phosphodiester bridge that links the glycerol backbone to some base, usually a nitrogenous one, such as choline, serine, or ethanolamine (Figures 10.4 and 10.5). The most abundant phospholipids in human tissues are phosphatidylcholine (also called lecithin), phosphatidylethanolamine, and phosphatidylserine. Note that C-2 of the phospholipids represents an asymmetric center (Fig-

ure 10.6). At physiologic pH, phosphatidylcholine and phosphatidylethanolamine have no net charge and exist as dipolar zwitterions, whereas phosphatidylserine has a net charge of −1, causing it therefore to be an acidic phospholipid. Note that phosphatidylethanolamine (PTE) is related to phosphatidylcholine in that trimethylation of PTE produces lecithin. Most phospholipids contain more than one kind of fatty acid per molecule, so that a given class of phospholipids from any tissue actually represents a family of molecular species. Phosphatidylcholine (PTC) contains mostly palmitic acid (16 : 0) or stearic acid (18 : 0) in the sn-1 position and primarily the unsaturated 18-carbon fatty acids oleic, linoleic, or linolenic in the sn-2 position. Phosphatidylethanolamine has the same saturated fatty acids as PTC at the sn-1 position but contains more of the long-chain polyunsaturated fatty acids—namely, 18 : 2, 20 : 4, and 22 : 6—at the sn-2 position.

Phosphatidylinositol is an acidic phospholipid that occurs in mammalian membranes (Figure 10.7). Phosphatidylinositol is rather unusual because it often contains almost exclusively stearic acid in the sn-1 position and arachidonic acid (20 : 4) in the sn-2 position.

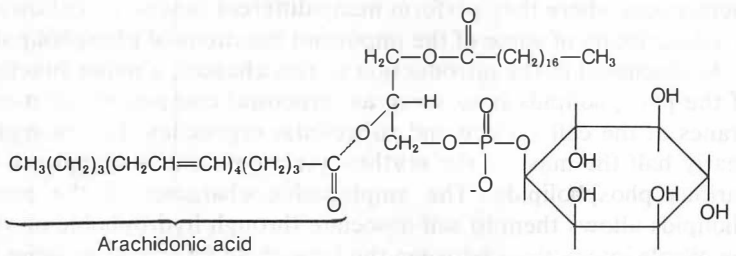


Figure 10.7
The structure of phosphatidylinositol.

Another phospholipid comprised of a polyol polar head group is phosphatidylglycerol. Phosphatidylglycerol (Figure 10.8) occurs in relatively large amounts in mitochondrial membranes and is a precursor of cardiolipin. Phosphatidylglycerol and phosphatidylinositol both carry a formal charge of −1 at neutral pH and are therefore acidic lipids.

The phospholipids that have been discussed so far contain only *O*-acyl residues attached to glycerol. *O*-(1-Alkenyl) substituents occur at C-1 of the sn-glycerol moiety of phosphoglycerides in combination with an *O*-acyl residue esterified to the C-2 position; compounds in this class are known as plasmalogens and are derivatives of plasmenic acid (Figure 10.9). Relatively large amounts of

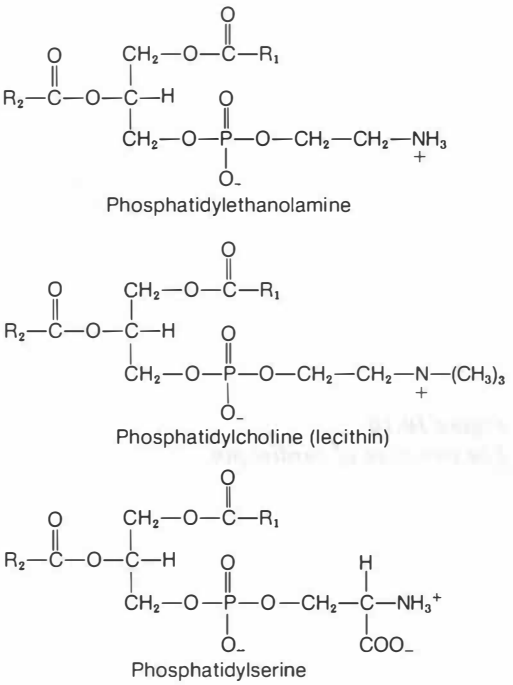


Figure 10.6
The structures of some common phospholipids.

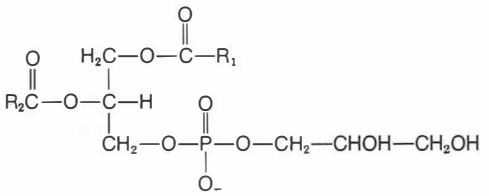


Figure 10.8
The structure of phosphatidylglycerol.

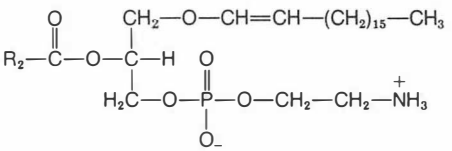


Figure 10.9
The structure of ethanolamine plasmalogen.

ethanolamine plasmalogen (also called plasmenylethanolamine) occur in myelin with lesser amounts in heart muscle where choline plasmalogen is abundant.

Cardiolipin, a very acidic (charge, −2) phospholipid, is composed of two molecules of phosphatidic acid linked together covalently through a molecule of glycerol. Cardiolipin is found primarily in the inner membrane of mitochondria and in bacterial membranes (Figure 10.10).

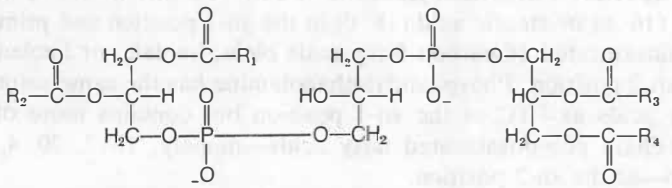


Figure 10.10
The structure of cardiolipin.

Functions of Phospholipids

Although present in body fluids such as plasma and bile, the phospholipids are found in highest concentration in the various cellular membranes where they perform many different functions. Following is a description of some of the important functions of phospholipids.

As discussed in the introduction to this chapter, a major function of the phospholipids is to serve as structural components of membranes of the cell surface and subcellular organelles. For example, nearly half the mass of the erythrocyte membrane is comprised of various phospholipids. The amphipathic character of the phospholipids allows them to self-associate through hydrophobic or van der Waals interactions between the long-chain fatty acyl moieties in adjacent molecules in such a way that the polar head groups project outward toward water where they can interact with protein molecules.

Phospholipids also play a role in activating certain enzymes. β-Hydroxybutyrate dehydrogenase is a mitochondrial enzyme imbedded in the inner membrane of that organelle, and it catalyzes the reversible abstraction of electrons from the ketone body, β-hydroxybutyrate (see page 476). The enzyme has an absolute requirement for phosphatidylcholine: removal of this phospholipid by organic solvent extraction renders the enzyme incapable of catalyzing the oxidation of β-hydroxybutyrate. Reconstitution of the dehydrogenase with other phospholipids such as phosphatidylserine and phosphatidylethanolamine cannot substitute for phosphatidylcholine in reactivating the enzyme.

Normal lung function depends upon a constant supply of an unusual phospholipid called dipalmitoyllecithin in which the lecithin

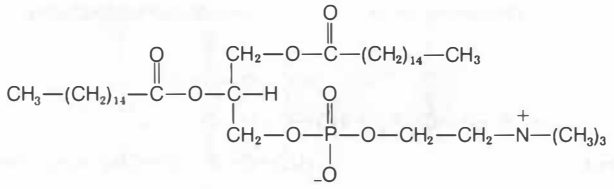


Figure 10.11
The structure of dipalmitoyllecithin.

molecule contains palmitic acid (16:0) residues in both the sn-1 and sn-2 positions. More than 80% of the phospholipid in the extracellular liquid layer that lines alveoli of normal lungs is contributed by dipalmitoyllecithin. This particular phospholipid—called surfactant—is produced by type II epithelial cells and prevents atelectasis at the end of the expiration phase of breathing (Figure 10.11). This lipid derives its name from its powerful capacity to decrease the surface tension of the aqueous surface layer of the lung (Figure 10.12). Lecithin molecules that do not contain two residues of palmitic acid are not effective in lowering the surface tension of the fluid layer lining alveoli.

During the third trimester—prior to the twenty-eighth week of gestation—the fetal lung is synthesizing primarily sphingomyelin. Normally, at this time, glycogen that has been stored in epithelial type II cells is converted to fatty acids and then to dipalmitoyllecithin. During lung maturation there is a good correlation between the increase in lamellar inclusion bodies that represent the intracellular pulmonary surfactant (phosphatidylcholine) storage organelles, called lamellar bodies, and the simultaneous decrease in glycogen content of type II pneumocytes. At the twenty-fourth week of gestation the type II granular pneumocytes appear in the alveolar epithelium, and within a few days they produce their typical osmiophilic lamellar inclusion bodies. The number of type II cells increases until the 32nd week at which time surface active agent appears in the lung and amniotic fluid. Surface tension decreases when the inclusion bodies increase in the type II cells. In the few

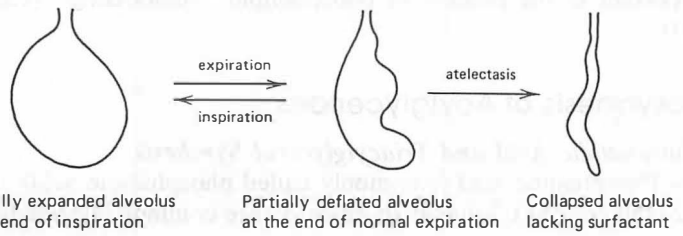


Figure 10.12
The role of surfactant in preventing atelectasis.

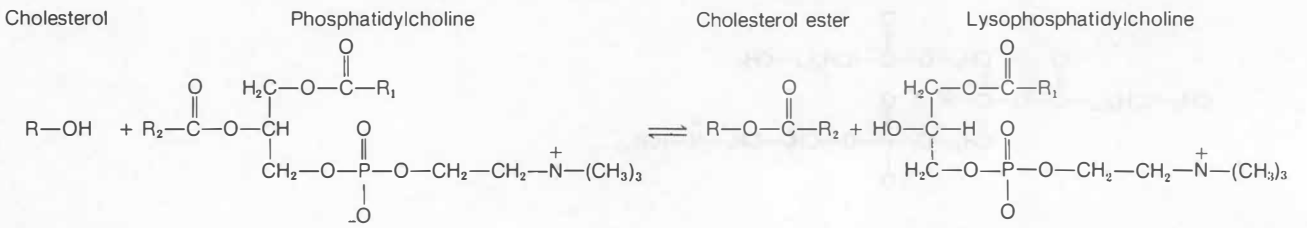


Figure 10.13
The lecithin : cholesterol acyltransferase (LCAT) reaction, where R—OH = cholesterol.

weeks before term one can perform screening tests on amniotic fluid to detect newborns that are at risk for RDS (Clin. Corr. 10.1). These tests are useful in timing elective deliveries, in applying vigorous preventive therapy to the newborn infant and to determine if the mother should be treated with a glucocorticoid drug such as betamethasone to accelerate maturation of the fetal lung.

Respiratory failure due to an insufficiency in surfactant can also occur in adults whose type II cells or surfactant-producing pneumocytes have been destroyed as an adverse side effect of the use of immunosuppressive medications or chemotherapeutic drugs.

The detergent properties of the phospholipids play an important role in bile where they function to solubilize cholesterol. An impairment in phospholipid production and secretion into bile can result in the formation of cholesterol and bile pigment gallstones.

Phospholipids are integral components of plasma lipoproteins and act to facilitate the solubilization of cholesterol, cholesterol esters, and triglycerides in chylomicrons, VLDL, and LDL. If phospholipid biosynthesis is impaired in the liver, then lipoproteins are not secreted and a fatty liver may result.

The fatty acids that comprise phospholipids can be transferred to various compounds, such as cholesterol. Such a transacylation process occurs by means of the LCAT (lecithin : cholesterol acyltranferase) reaction shown in Figure 10.13 in which the fatty acid in the sn-2 position of phosphatidylcholine is transferred to the 3-hydroxyl group of free cholesterol (R—OH) forming cholesterol ester and lysophosphatidylcholine. A specific lipoprotein—namely, HDL—is the actual substrate for LCAT. The cholesterol esters produced in this reaction are necessary for maintenance of the normal structure of plasma lipoproteins. Exchange reactions of this type are also important in the process of phospholipid "remodeling" (see page 501).

CLIN. CORR. 10.1 RESPIRATORY DISTRESS SYNDROME

The respiratory distress syndrome (RDS) is a major cause of neonatal morbidity and mortality in many countries. It accounts for approximately 15–20% of all neonatal deaths in Western countries and somewhat less in the developing countries. The disease affects only premature babies and its incidence varies directly with the degree of prematurity. Premature babies develop RDS because of immaturity of their lungs, resulting from a deficiency of pulmonary surfactant. The maturity of the fetal lung

Biosynthesis of Acylglycerides

Phosphatidic Acid and Triacylglycerol Synthesis

L-α-Phosphatidic acid (commonly called phosphatidic acid) and sn-1,2-diglyceride (1,2-diacyl-sn-glycerol) are common intermediates in

the pathways of phospholipid and triacylglycerol biosynthesis. Furthermore, the biosynthesis of triacylglycerides proceeds by way of a pathway comprised of enzymes shared by the pathway of phospholipid synthesis. Essentially all cells are capable of synthesizing phospholipids to some degree (except mature erythrocytes), whereas triacylglyceride biosynthesis occurs only in liver, adipose tissue, and intestine. There are two pathways for phosphatidic acid biosynthesis in mammals: one begins with the acylation of α-glycerophosphate and the other with the acylation of dihydroxyacetone phosphate. In most tissues the principal pathway for phosphatidic acid synthesis begins with α-glycerophosphate (sn-glycerol 3-phosphate), and there are two sources of this triose phosphate. The most general source of α-glycerophosphate, particularly in adipose tissue, is from reduction of the glycolytic intermediate, dihydroxyacetone phosphate, in the reaction catalyzed by α-glycerophosphate dehydrogenase:

$$\text{Dihydroxyacetone phosphate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{glycerol 3-phosphate} + \text{NAD}^+$$

A few specialized tissues, including the liver, kidney, and intestine, derive α-glycerophosphate by means of the glycerol kinase reaction:

$$\text{Glycerol} + \text{ATP} \xrightarrow{\text{Mg}^{2+}} \text{glycerol 3-phosphate} + \text{ADP}$$

Adipocytes are devoid of glycerol kinase and must rely upon α-glycerophosphate dehydrogenase as a source of glycerol 3-phosphate. When triglycerides are broken down in the fat cell, the glycerol that is generated cannot be utilized by the fat cell. Glycerol diffuses from the fat cell and is transported to the liver, where it is metabolized by the glycerol kinase reaction.

The next two steps in phosphatidic acid biosynthesis involve stepwise transfer of long-chain fatty acyl groups from the activated donor, fatty acyl CoA. The first acyltransferase (I) is called glycerol phosphate : acyltransferase and attaches predominantly saturated fatty acids and oleic acid to the sn-1 position to produce 1-acylglycerol phosphate or α-lysophosphatidic acid. The second enzyme (II), 1-acylglycerol phosphate : acyltransferase, catalyzes the acylation of the sn-2 position, usually with an unsaturated fatty acid (Figure 10.14). The high energy, highly reactive donor of acyl groups is the coenzyme A thioester derivative of the long-chain fatty acids.

The specificity of the two acyltransferases does not always match the fatty acid asymmetry that one finds in the phospholipids of a particular cell. Remodeling reactions to be discussed below function

can be predicted antenatally by measuring the lecithin/sphingomyelin (L/S) ratio in the amniotic fluid. The mean L/S ratio in normal pregnancies increases gradually with gestation until about 31 or 32 weeks when the slope rises sharply. The ratio of 2.0 that is characteristic of the term infant at birth is achieved at the gestational age of about 34 weeks. In terms of predicting pulmonary maturity, the critical L/S ratio is 2.0 or greater. The risk of developing RDS when the L/S ratio is less than 2.0 has been worked out: for an L/S ratio of 1.5–1.9, the risk is approximately 40%, and for a ratio less than 1.5 the calculated risk of developing RDS is about 75%.

Although the L/S ratio in amniotic fluid is still widely used to predict the risk of RDS, the results are unrealiable if the amniotic fluid specimen has been contaminated by blood or meconium obtained during a complicated pregnancy. In recent years the determination of saturated palmitoylcholine (SPC) has been found to be more specific and a more sensitive predictor of the RDS than the L/S ratio. There are rapid techniques for the isolation of saturated phosphatidylcholine that can be used to assay the surfactant lecithin content in amniotic fluid.

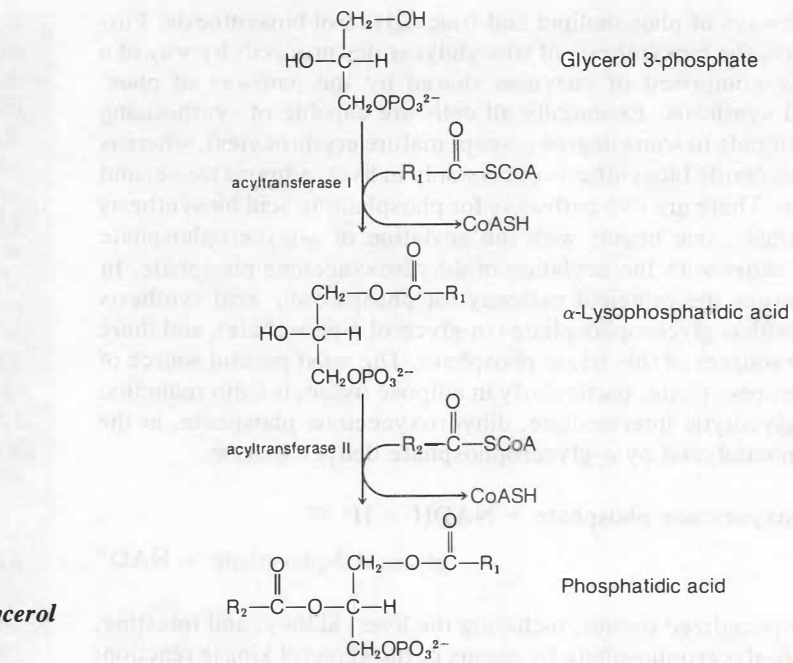


Figure 10.14
Phosphatidic acid biosynthesis from glycerol 3-phosphate.

to modify the fatty acid composition at the C-1 and C-2 positions of the glycerol phosphate backbone.

Cytoplasmic phosphatidic acid phosphatase (also called phosphatidic acid phosphohydrolase) hydrolyzes phosphatidic acid (1,2-diacylglycerophosphate) that is generated on the endoplasmic reticulum, thereby yielding sn-1,2-diacylglycerol (diglyceride) that serves as the branch point in triglyceride and phospholipid synthesis (Figure 10.15).

Phosphatidic acid can also be formed by a second pathway that begins with dihydroxyacetone phosphate (DHAP). The DHAP

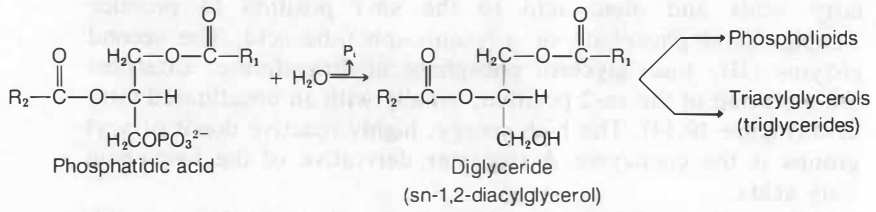


Figure 10.15
The phosphatidic acid phosphatase reaction.

pathway is usually an alternative supportive route employed by some tissues to produce phosphatidic acid. DHAP, an intermediate in glycolysis, is first acylated with a long-chain fatty acid residue from the corresponding fatty acyl CoA thioester, thereby producing 1-acyldihydroxyacetone phosphate:

dihydroxyacetone phosphate + fatty acyl CoA $\rightleftharpoons$

1-acyldihydroxyacetone phosphate + CoASH

The enzyme that catalyzes this reaction is named acyl CoA : dihydroxyacetone phosphate acyltransferase, and it is located in peroxisomes (i.e., microbodies). In the next step, reduction of the keto group in the sn-2 position by an NADPH-dependent dehydrogenase generates 1-acylglycerophosphate:

1-acyldihydroxyacetone phosphate + NADPH + H^+ $\rightleftharpoons$

1-acylglycerophosphate + $NADP^+$

Subsequent acylation of 1-acylglycerophosphate with a molecule of long-chain fatty acid yields diacylglycerophosphate (phosphatidic acid). The DHAP and α-glycerophosphate pathways for phosphatidic acid biosynthesis converge at the structure sn-1 acylglycerophosphate (1-acyl-sn-glycero-3-phosphate).

The last step in triglyceride biosynthesis requires acylation of the sn-3 position in 1,2-diacylglycerol; this reaction is catalyzed by the enzyme diglyceride acyltransferase:

Diglyceride + fatty acyl CoA $\rightarrow$ triacylglycerol + CoASH

Biosynthesis of Specific Phospholipids

Two different pathways exist for the biosynthesis of *phosphatidylcholine* (lecithin), and they are distinguished by whether activation involves the phosphocholine polar head group, or the diglyceride moiety; in either case activation involves the high energy compound, CTP. In the first case, activation of choline is accomplished by the following reactions. First, free choline is phosphorylated by ATP in a reaction catalyzed by choline kinase (Figure 10.16). Phosphocholine in turn is converted to CDP-choline at the expense of CTP in the reaction catalyzed by phosphocholine cytidyltransferase; note that inorganic pyrophosphate (PP_i) is a product of this reaction resulting from attack by the phosphoryl residue of phosphocholine on the internal α-phosphorus atom of CTP. The high energy pyrophosphoryl bond in CDP-choline is very unstable and reactive such that the phosphocholine moiety can be transferred readily to the nucleo-

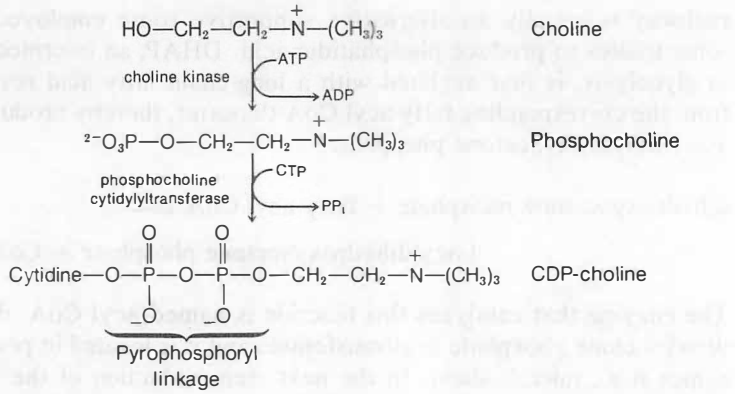


Figure 10.16
The biosynthesis of CDP-choline from choline.

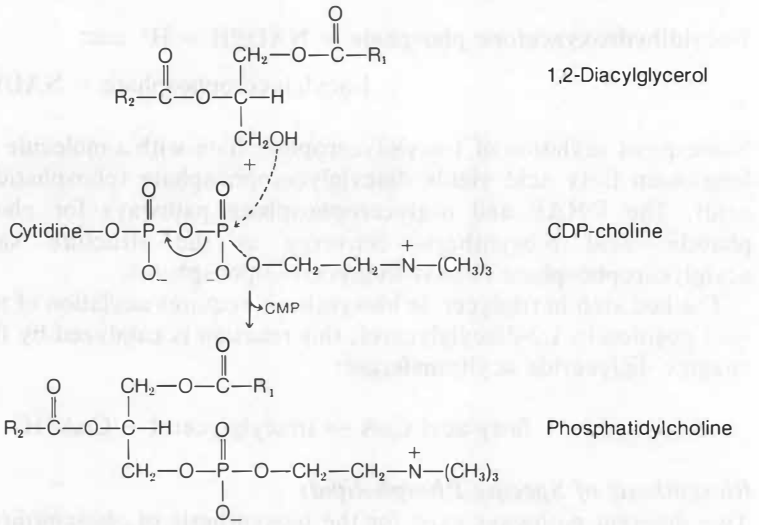


Figure 10.17
The choline phosphotransferase reaction.

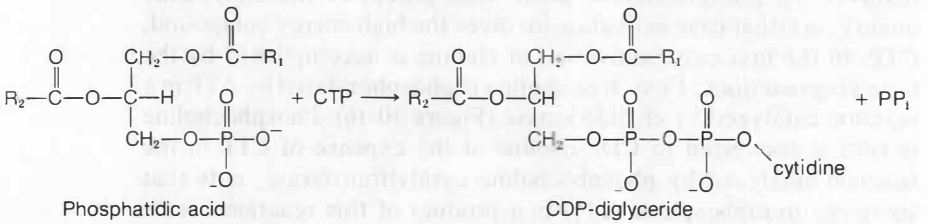


Figure 10.18
The biosynthesis of CDP-diglyceride by the CTP : phosphatidic acid cytidyltransferase reaction.

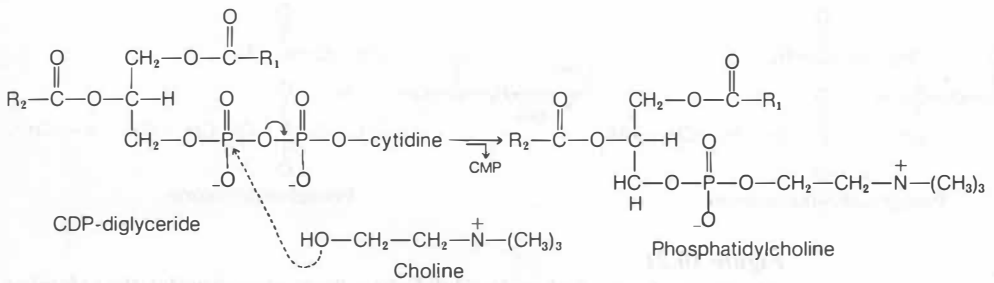


Figure 10.19
The biosynthesis of phosphatidylcholine from CDP-diglyceride and free choline.

philic center provided by the OH group at position 3 of 1,2-diglyceride in the reaction catalyzed by choline phosphotransferase as shown in Figure 10.17. This is the principal pathway for the synthesis of dipalmitoyllecithin in the lung. In an alternate pathway for the biosynthesis of phosphatidylcholine, CTP can be used instead to activate the diglyceride moiety of phosphatidic acid in the reaction catalyzed by CTP : phosphatidic acid cytidyltransferase shown in Figure 10.18. In this case, the activated pyrophosphoryl center occurs attached to the diglyceride moiety and the nucleophilic center in phosphatidylcholine synthesis is the —OH group of free choline as shown in Figure 10.19. The two pathways for phosphatidylcholine (PTC) synthesis are summarized in Figure 10.20.

Phosphatidylcholine can also be formed by repeated methylation of the phospholipid phosphatidylethanolamine. Phosphatidylethanolamine *N*-methyltransferase catalyzes the transfer of methyl groups—one at a time—from *S*-adenosylmethionine (SAM) to phosphatidylethanolamine to produce phosphatidylcholine (Figure 10.21).

The primary pathway for *phosphatidylethanolamine* synthesis in liver and brain involves the microsomal enzyme ethanolamine phosphotransferase that catalyzes the reaction shown in Figure 10.22. This enzyme is particularly abundant in liver. CDP-ethanolamine is formed through the reaction catalyzed by ethanolamine kinase:

$$\text{Ethanolamine} + \text{ATP} \xrightarrow[Mg^{2+}]{} \text{phosphorylethanolamine} + \text{ADP}$$

and the phosphorylethanolamine cytidylyltransferase reaction:

$$\text{Phosphorylethanolamine} + \text{CTP} \xrightarrow[Mg^{2+}]{} \text{CDP-ethanolamine} + PP_i$$

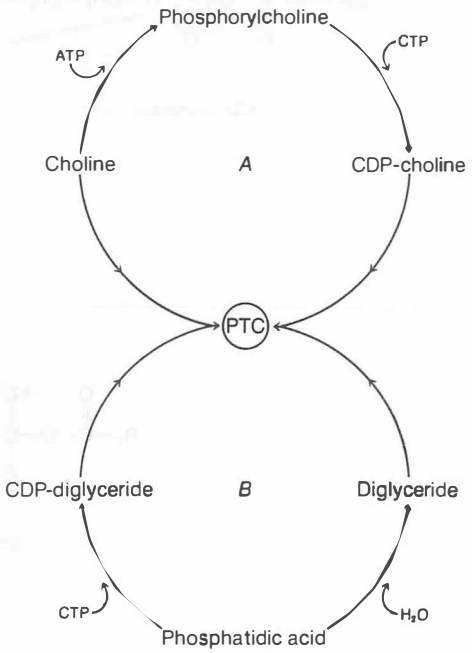


Figure 10.20
Summary of the two principal pathways for phosphatidylcholine (PTC) biosynthesis.
In pathway A, choline is activated by CTP, whereas in pathway B, the diglyceride moiety is activated by CTP.

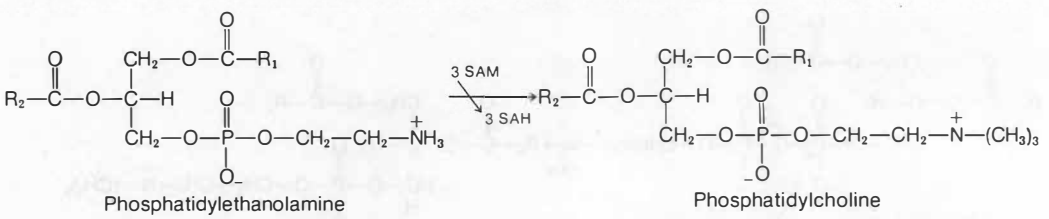


Figure 10.21
The biosynthesis of phosphatidylcholine from phosphatidylethanolamine and S-adenosylmethionine (SAM); SAH (S-adenosylhomocysteine).

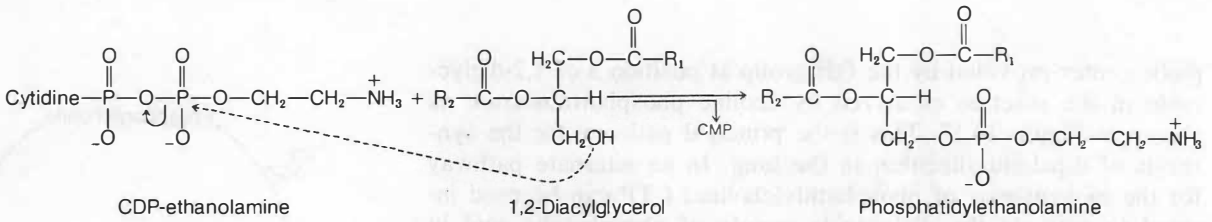


Figure 10.22
The biosynthesis of phosphatidylethanolamine from CDP-ethanolamine and diglyceride; the reaction is catalyzed by ethanolamine phosphotransferase.

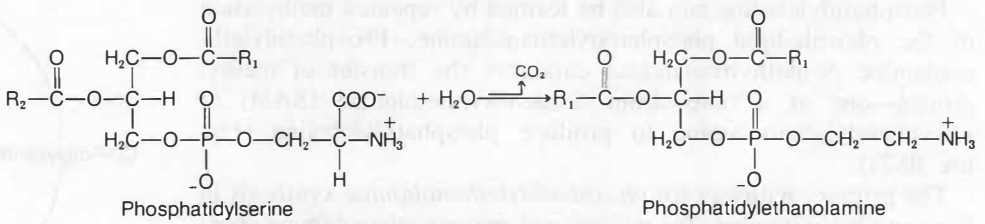


Figure 10.23
The formation of phosphatidylethanolamine by the decarboxylation of phosphatidylserine.

Liver mitochondria can also generate phosphatidylethanolamine by decarboxylation of phosphatidylserine; however, this is thought to represent only a minor pathway in phosphatidylethanolamine synthesis (Figure 10.23).

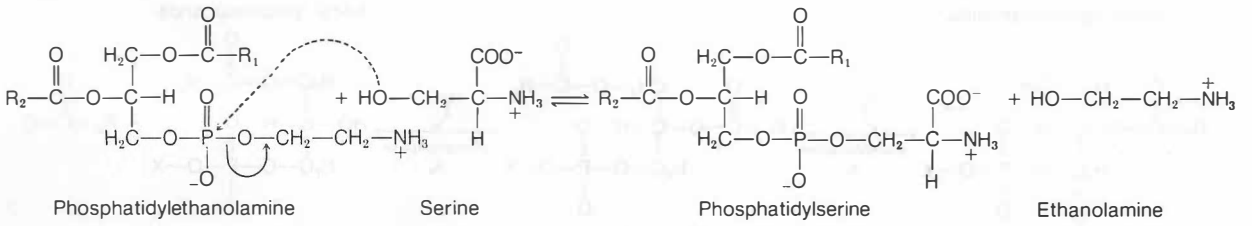


Figure 10.24
The biosynthesis of phosphatidylserine from serine and phosphatidylethanolamine by "base exchange."

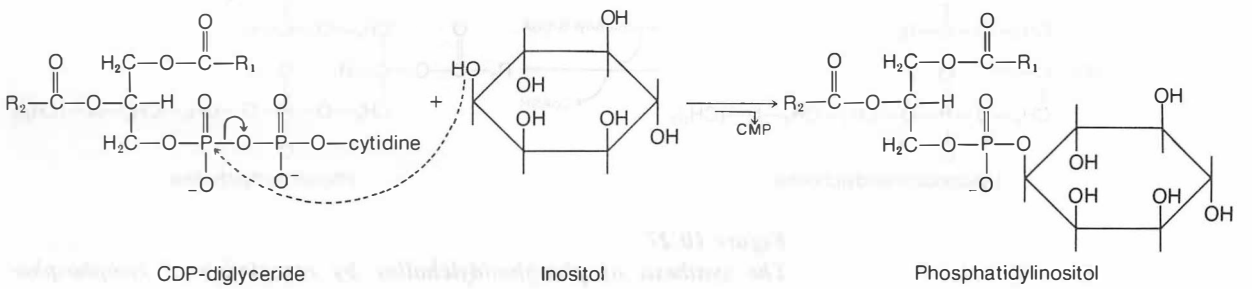


Figure 10.25
The biosynthesis of phosphatidylinositol.

The major source of *phosphatidylserine* appears to be provided by the "base-exchange" reaction shown in Figure 10.24, where the polar head group of phosphatidylethanolamine is exchanged for the amino acid serine; since there is no net change in the number or kinds of bonds, this reaction is reversible and has no requirement for ATP or any other high energy compound. The reaction is initiated by attack on the phosphodiester bond of phosphatidylethanolamine by the hydroxyl group of serine.

Phosphatidylinositol is made via CDP-diglyceride and free inositol (Figure 10.25).

Remodeling Reactions and the Role of Phospholipases in Phospholipid Synthesis

The activities of two phospholipases, phospholipase A_1 and phospholipase A_2, can be demonstrated in many tissues, and there is considerable evidence that each plays a role in the formation of

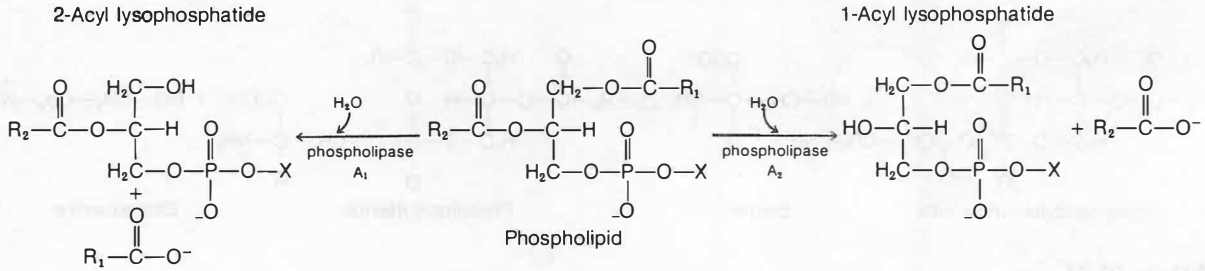


Figure 10.26
The reactions catalyzed by phospholipase A_1 and phospholipase A_2.

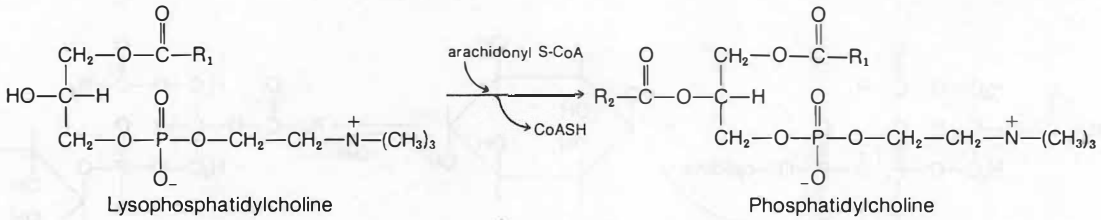


Figure 10.27
The synthesis of phosphatidylcholine by reacylation of lysophosphatidylcholine where $R_2—\overset{O}{\overset{\|}{C}}—O—$ represents arachidonic acid.

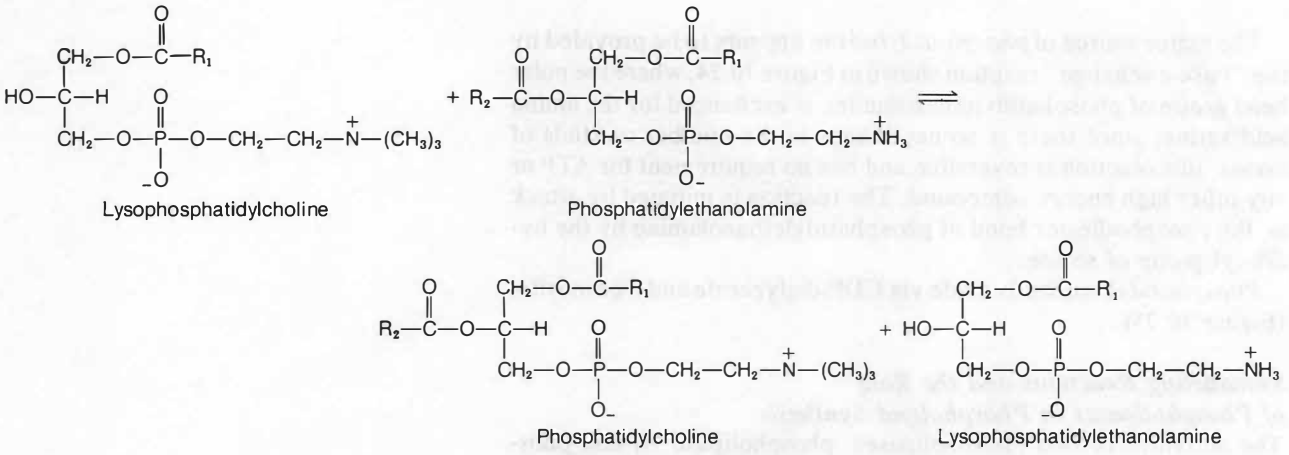


Figure 10.28
Formation of phosphatidylcholine by lysolecithin exchange, where $R_2—\overset{O}{\overset{\|}{C}}—O—$ represents arachidonic acid.

specific phospholipid structures containing the proper kinds of fatty acids in the sn-1 and sn-2 positions. It has been found that most of the fatty acyl CoA transferases and phospholipid synthesizing enzymes discussed above are lacking with regard to the specificity required to account for the asymmetric position or distribution of fatty acids found in many tissue phospholipids. That is, the fatty acids that are found in the sn-1 and sn-2 positions of the various phospholipids are often not the same ones that were transferred to the glycerol backbone in the initial acyl transferase reactions of the phospholipid biosynthesis pathways. Phospholipases A_1 and A_2 catalyze the reactions indicated in Figure 10.26 where X represents the polar head group of a phospholipid. The products of the action of phospholipases A_1 and A_2 are called lysophosphatides.

For example, if it becomes necessary for a cell to remove some undesired fatty acid, such as stearic acid from the sn-2 position of phosphatidylcholine, and replace it by a more unsaturated one like arachidonic acid, then this can be accomplished by the action of phospholipase A_2 followed by a reacylation step. The insertion of arachidonic acid into the 2 position of sn-2-lysophosphatidylcholine can then be accomplished by one of two means; either by direct acylation from arachidonyl CoA (Figure 10.27) or from some other arachidonic acid-containing phospholipid by an exchange-type reaction (Figure 10.28). Remodeling by direct acylation would be accomplished by an arachidonic acid-specific acyl CoA transacylase. The lysolecithin exchange reaction is catalyzed by lysolecithin: lecithin acyltransferase (LLAT) (Figure 10.28). Note that, since there is no change in either the number or nature of the bonds involved in products and reactants, there is no ATP requirement for this acylation reaction.

Lysophospholipids, particularly sn-1-lysophosphatidylcholine, can also serve as sources of fatty acid in the remodeling reactions. For example, the following diagram summarizes the remodeling reactions that might be involved in the synthesis of dipalmitoyllecithin (surfactant) from 1-palmitoyl-2-oleoylphosphatidylcholine. Note that sn-1-palmitoyl lysolecithin is the source of palmitic acid in the acyltransferase exchange reaction (Figure 10.29).

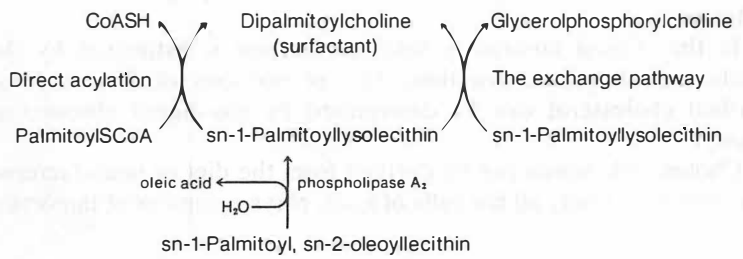


Figure 10.29
Two pathways for the biosynthesis of dipalmitoyllecithin from sn-1 palmitoyllysolecithin.

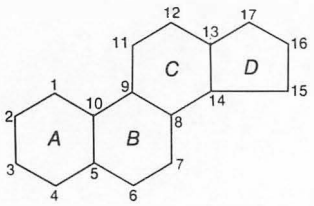


Figure 10.30
The cyclopentanophenanthrene ring.

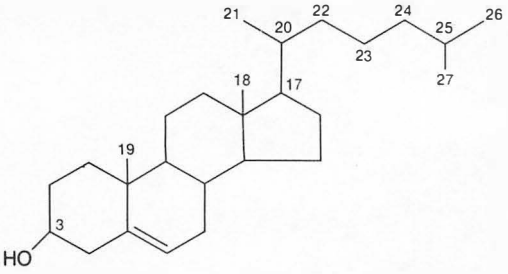


Figure 10.31
The structure of cholesterol (cholest-5-en-3β-ol).

10.3 CHOLESTEROL

Structure and Function of Cholesterol

Cholesterol is an alicyclic compound whose structure includes (1) the cyclopentanophenanthrene nucleus with its four fused rings, (2) a single hydroxyl group at C-3, (3) an unsaturated center between carbon atoms 5 and 6, (4) an eight-membered branched hydrocarbon chain attached to the D ring at position 17, and (5) a methyl group (designated C-19) attached at position 10 and another methyl group (designated C-18) attached at position 13. (See Figures 10.30 and 10.31).

In terms of physical properties, cholesterol is a lipid with very low solubility in water: at 25°C, the limit of solubility is approximately 0.2 mg/100 ml, or 4.7 μM. The actual concentration of cholesterol in plasma of healthy people is usually 150–200 mg/dl: on a milligram basis, this value is almost twice the normal concentration of blood glucose. The very high solubility of cholesterol in blood is due to the presence of proteins called plasma lipoproteins (mainly LDL and VLDL) that have the ability to bind and thereby solubilize large amounts of cholesterol (see page 464).

Actually, only about 30% of the total circulating cholesterol occurs free as such; approximately 70% of the cholesterol in plasma lipoproteins exists in the form of cholesterol esters where some long-chain fatty acid, usually linoleic acid, is attached by an ester bond to the OH—group on carbon-3 of the A ring. The presence of the long-chain fatty acid residue enhances the hydrophobicity of cholesterol (Figure 10.32).

Cholesterol is also abundant in bile where the normal concentration is 390 mg/100 ml. In contrast to the finding of predominantly cholesterol esters in plasma, only 4% of the cholesterol in bile is esterified to some long-chain fatty acid. Bile does not contain appreciable amounts of any of the lipoproteins and the solubilization of free cholesterol is achieved in part by the detergent property of phospholipids present in bile that are derived from the liver. A chronic disturbance in phospholipid metabolism in the liver can result in the deposition of cholesterol-rich gallstones. Bile salts, which are derivatives of cholesterol, also aid in keeping cholesterol in solution.

In the clinical laboratory total cholesterol is estimated by the Lieberman–Burchard reaction. The proportions of free and esterified cholesterol can be determined by gas–liquid chromatography.

Cholesterol, which can be derived from the diet or manufactured de novo in virtually all the cells of man, plays a number of important

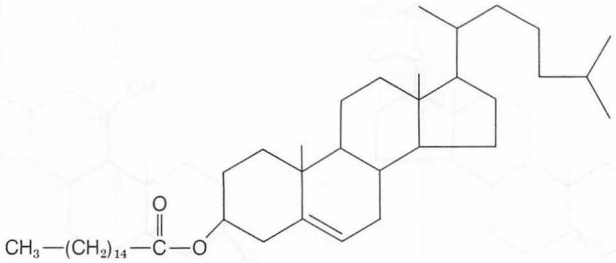


Figure 10.32
The structure of cholesterol (palmitoyl-) ester.

roles. It is the major sterol in man and a component of virtually all cell surfaces and intracellular membranes. Cholesterol is especially abundant in the myelinated structures of the brain and central nervous system but is present in small amounts in the inner membrane of the mitochondrion. In contrast to the situation in plasma, most of the cholesterol in cellular membranes occurs in the free, unesterified form.

The second role of cholesterol is that it is the immediate precursor of the bile acids that are synthesized in the liver and which function to facilitate the absorption of dietary triacylglycerols and fat-soluble vitamins (Chapter 24). It is important to realize that the ring structure of cholesterol cannot be metabolized to CO_2 and water in man. The route of excretion of cholesterol is by the way of the liver and gallbladder through the intestine in the form of bile acids.

The third physiological role of cholesterol is that it is the precursor of the various steroid hormones. Progesterone is the 21-carbon keto steroid sex hormone secreted by the corpus luteum of the ovary. Pregnane is the nuclear parent of progesterone (Figures 10.33 and 10.34). The metabolically powerful corticosteroids of the adrenal cortex are derived from cholesterol. The addition of an hydroxyl group at C-21 of progesterone produces deoxycorticosterone; 11-

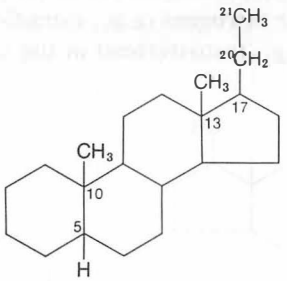


Figure 10.33
The structure of pregnane.

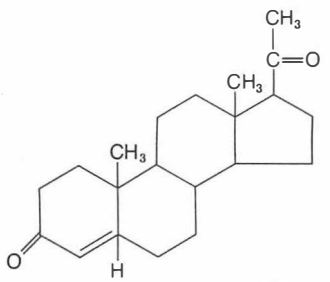


Figure 10.34
The structure of progesterone.

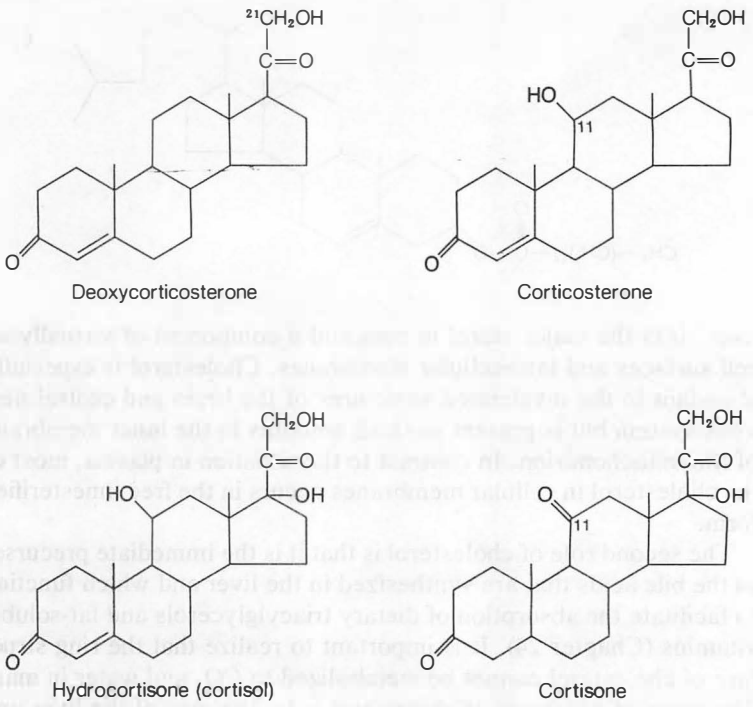


Figure 10.35
The steroid hormones of the adrenal cortex.

hydroxylation of deoxycorticosterone in turn generates corticosterone; and subsequent 17-hydroxylation yields hydrocortisone (cortisol). Oxidation of hydrocortisone at C-11 gives cortisone (Figure 10.35). The mineralocorticoid aldosterone is derived from cholesterol in the zona glomerulosa tissue of the cortex of the adrenal gland (Figure 10.36). Cholesterol also serves as the precursor to the female steroid hormones, the estrogens (e.g., estradiol), in the ovary and to the male steroids (e.g., testosterone) in the testes (Figure 10.37).

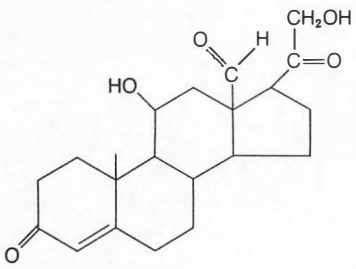


Figure 10.36
The structure of aldosterone.

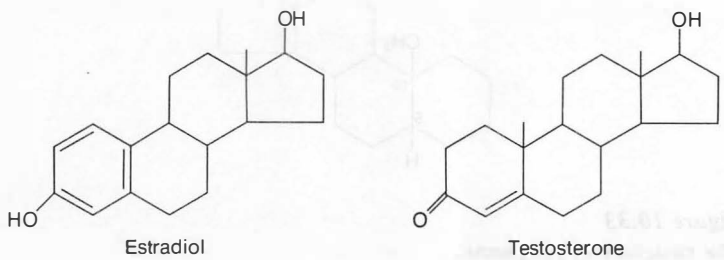


Figure 10.37
The structure of the estrogen estradiol and the androgen testosterone.

It is noteworthy that, although all of the steroid hormones are structurally related to and biochemically derived from cholesterol, they have widely different physiological properties that control spermatogenesis, pregnancy, lactation and parturition, mineral balance, and energy (carbohydrate and fat) metabolism. The metabolism and function of the steroid hormones will be discussed in Chapter 15.

The fundamental hydrocarbon skeleton of cholesterol is also found in the plant sterols, ergosterol, a precursor to vitamin D, and 7-dehydrocholesterol (Figure 10.38). After ingestion, ergosterol and 7-dehydrocholesterol are converted in the skin by ultraviolet irradiation to vitamin D_3 (cholecalciferol). Vitamin D_3 is involved in calcium and phosphorus metabolism (see Figure 10.39).

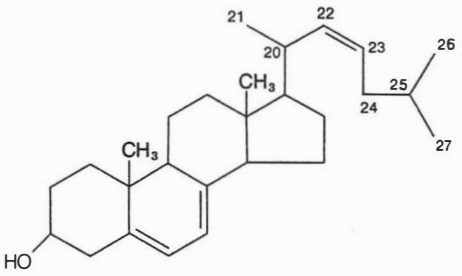
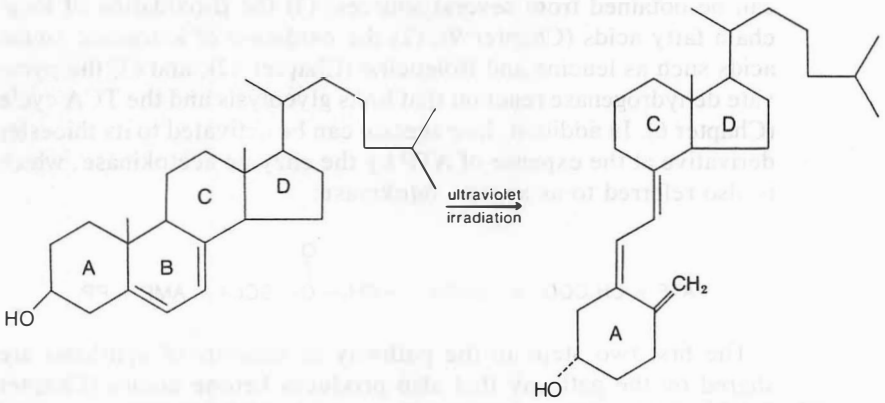


Figure 10.38
The structure of ergosterol.



7-Dehydrocholesterol Cholecalciferol (vitamin D_3)

Figure 10.39
The conversion of 7-dehydrocholesterol to cholecalciferol.

Cholesterol Biosynthesis

Although the de novo biosynthesis of cholesterol occurs in virtually all cells, this capacity is greatest in certain tissues, particularly the liver, intestine, adrenal cortex, and reproductive tissues, including ovaries, testes, and placenta. From an inspection of its structure it is apparent that cholesterol biosynthesis will require a source of carbon atoms and considerable reducing power to generate the numerous carbon–hydrogen and carbon–carbon bonds. All of the carbon atoms of cholesterol are derived from acetate. Reducing power in the form of NADPH is provided by enzymes of the hexose monophosphate shunt, specifically, glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase. It should be remembered that

for each glucose molecule oxidized by way of the hexose monophosphate shunt, 2 equiv of NADPH are produced. With regard to the requirement for high energy bond-containing compounds, the pathway of cholesterol synthesis is driven in large part by the hydrolysis of the high energy thioester bonds of acetyl CoA and the high energy phosphoanhydride bonds of ATP. Realizing that cholesterol biosynthesis occurs in the cytoplasm of cells, let us consider the individual steps in the pathway.

Formation of Mevalonic Acid from Acetate

The first compound unique to the pathway of cholesterol biosynthesis is mevalonic acid. Mevalonic acid is derived from the two-carbon precursor acetyl CoA that is located at the hub of the pathways of fat, carbohydrate, and amino acid metabolism. Acetyl CoA can be obtained from several sources: (1) the β-oxidation of long-chain fatty acids (Chapter 9); (2) the oxidation of ketogenic amino acids such as leucine and isoleucine (Chapter 12); and (3) the pyruvate dehydrogenase reaction that links glycolysis and the TCA cycle (Chapter 6). In addition, free acetate can be activated to its thioester derivative at the expense of ATP by the enzyme acetokinase, which is also referred to as acetate thiokinase:

$$\text{ATP} + \text{CH}_3\text{COO}^- + \text{CoASH} \longrightarrow \text{CH}_3\text{—}\overset{\overset{\text{O}}{\|}}{\text{C}}\text{—SCoA} + \text{AMP} + \text{PP}_i$$

The first two steps in the pathway of cholesterol synthesis are shared by the pathway that also produces ketone bodies (Chapter 9). Two molecules of acetyl CoA condense to form acetoacetyl CoA in a reaction catalyzed by acetoacetyl CoA thiolase (acetyl CoA : acetyl CoA acetyltransferase):

$$\text{CH}_3\text{—}\overset{\overset{\text{O}}{\|}}{\text{C}}\text{—SCoA} + \text{CH}_3\text{—}\overset{\overset{\text{O}}{\|}}{\text{C}}\text{—SCoA} \longrightarrow \text{CH}_3\text{—}\overset{\overset{\text{O}}{\|}}{\text{C}}\text{—CH}_2\text{—}\overset{\overset{\text{O}}{\|}}{\text{C}}\text{—SCoA} + \text{CoASH}$$

Note that the formation of the carbon–carbon bond in acetoacetyl CoA in this reaction is favored energetically by the cleavage of a thioester bond and the generation of free coenzyme A (CoASH).

The next step introduces a third molecule of acetyl CoA into the cholesterol pathway and forms the branched-chain compound 3-hydroxy-3-methylglutaryl CoA (HMG CoA) (Figure 10.40). This condensation reaction is catalyzed by HMG CoA synthase (3-hydroxy-3-methylglutaryl CoA : acetoacetyl CoA lyase). Liver parenchymal cells contain two isoenzyme forms of HMG CoA synthase; one is found in the cytoplasm and is involved in cholesterol synthesis, while the other has a mitochondrial location and

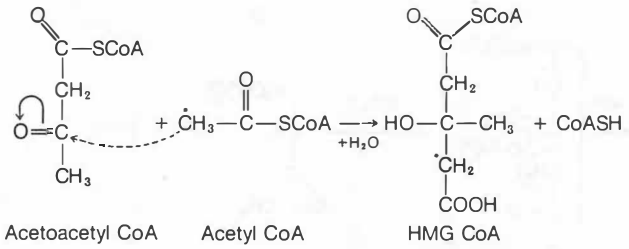


Figure 10.40
The HMG CoA synthase reaction.

functions in the pathway that forms ketone bodies. In the HMG CoA synthase reaction, an aldol condensation occurs between the methyl carbon of acetyl CoA and the β-carbonyl group of acetoacetyl CoA with the simultaneous hydrolysis of the thioester bond of acetyl CoA. Note that the thioester bond in the original acetoacetyl CoA substrate molecule remains intact.

HMG CoA can also be formed from the oxidative degradation of the branched-chain amino acid leucine, which proceeds through the intermediates 3-methylcrotonyl CoA and 3-methylglutaconyl CoA (Chapter 12).

The step that produces the unique compound mevalonic acid from HMG CoA is catalyzed by the important microsomal enzyme HMG CoA reductase (mevalonate : $NADP^+$ oxidoreductase) that has an absolute requirement for NADPH as the reductant (Figure 10.41). Note that this reductive step (1) consumes two molecules of NADPH from the pentose phosphate pathway, (2) results in the hydrolysis of the thioester bond of HMG CoA, and (3) generates a primary alcohol residue in mevalonate. This reduction reaction is irreversible and produces R-(+) mevalonate, which contains six carbon atoms.

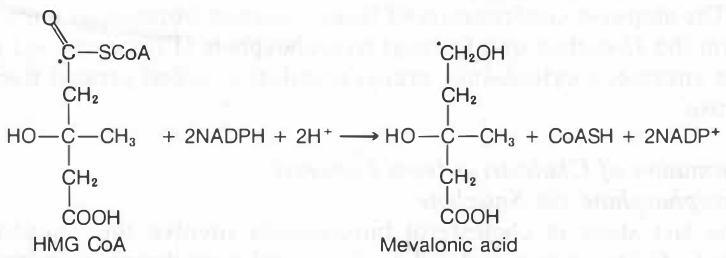


Figure 10.41
The HMG CoA reductase reaction.

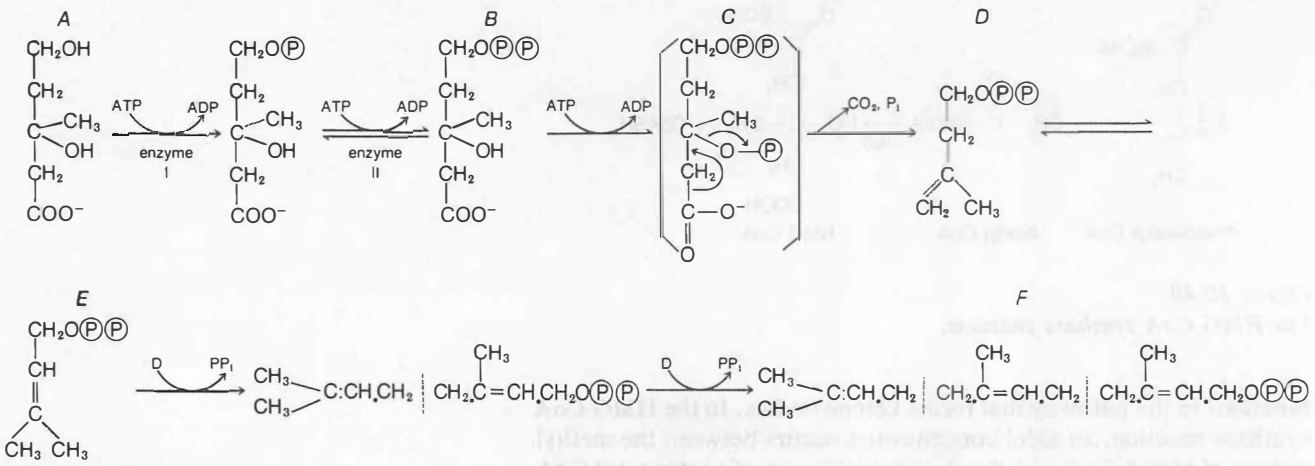


Figure 10.42
The formation of farnesyl-PP (F) from mevalonate (A).
The dotted lines divide the molecules into isoprenoid-derived units.

Conversion of Mevalonic Acid to Farnesyl Pyrophosphate

The various reactions involved in the conversion of mevalonate to farnesyl pyrophosphate are described below and summarized in Figure 10.42.

The stepwise transfer of the terminal γ-phosphate group from two molecules of ATP to mevalonate (A) to form 5-pyrophosphomevalonate (B) are catalyzed by mevalonate kinase (enzyme I) and phosphomevalonate kinase (enzyme II). The next step effects the decarboxylation of 5-phosphomevalonate and generates Δ^3-isopentenyl pyrophosphate (D); this reaction is catalyzed by pyrophosphomevalonate decarboxylase. In this ATP-dependent reaction in which ADP, P_i, and CO_2 are produced it is thought that decarboxylation–dehydration proceeds by way of the triphosphate intermediate, 3-phosphomevalonate 5-pyrophosphate (C). Next, isopentenyl pyrophosphate is converted to its allylic isomer 3,3-dimethylallyl pyrophosphate (E) in a reversible reaction catalyzed by isopentenyl pyrophosphate isomerase.

The stepwise condensation of three 5-carbon isopentenyl units to form the 15-carbon unit farnesyl pyrophosphate (F) is catalyzed by one enzyme, a cytoplasmic prenyl transferase called geranyl transferase.

Formation of Cholesterol from Farnesyl Pyrophosphate via Squalene

The last steps in cholesterol biosynthesis involve the "head-to-head" fusion of two molecules of farnesyl pyrophosphate to form squalene and finally the cyclization of squalene to yield cholesterol. The reaction that produces the 30-carbon molecule of squalene from

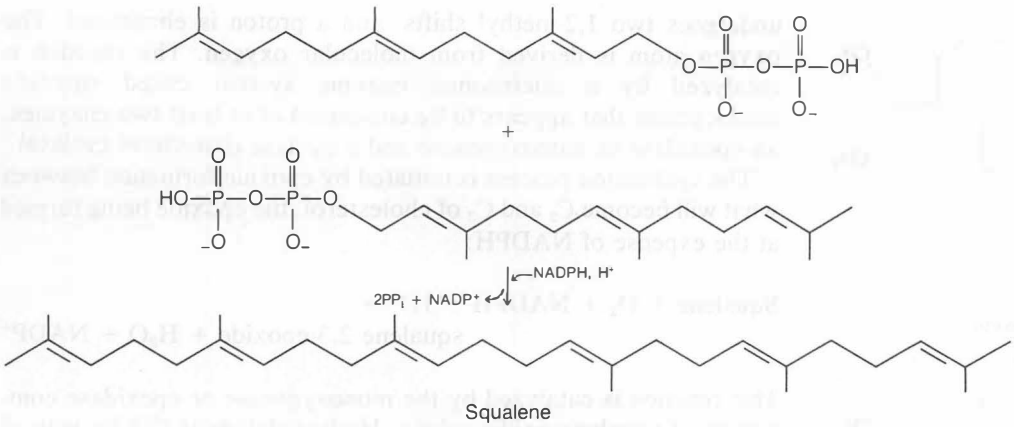


Figure 10.43
The formation of squalene from two molecules of farnesyl pyrophosphate.

two 15-carbon farnesyl pyrophosphate moieties is unlike the previous carbon–carbon bond-forming reactions in the pathway (Figure 10.43).

In this reaction catalyzed by the microsomal enzyme squalene synthetase, two pyrophosphate groups are released, with loss of a hydrogen atom from one molecule of farnesyl pyrophosphate and replacement by a hydrogen from NADPH. Several different intermediates probably occur between farnesyl pyrophosphate and squalene. By rotation about carbon–carbon single bonds, the conformation of squalene indicated in Figure 10.44 can be obtained. Note the similarity of the overall shape of the compound to cholesterol. Observe also that squalene is devoid of oxygen atoms.

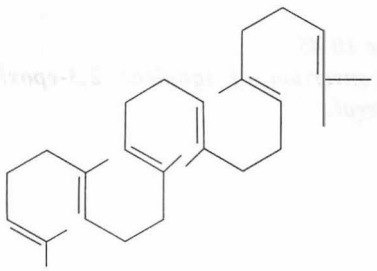


Figure 10.44
The structure of squalene, C_{30}.

Cholesterol biosynthesis from squalene proceeds through the intermediate lanosterol, which contains the fused tetracyclic ring system and an 8-carbon side chain:

Squalene → squalene 2,3-epoxide → lanosterol

The many carbon–carbon bonds that are formed during the cyclization of squalene are generated in a concerted fashion as indicated in Figure 10.45. Note that the —OH group of lanosterol projects above the plane of the A ring; this is referred to as the β orientation. Groups that extend down below the ring in a trans relationship to the —OH group are designated as α (alpha) by a dotted line. During this reaction sequence a hydroxyl group is added to C-3, a methyl group

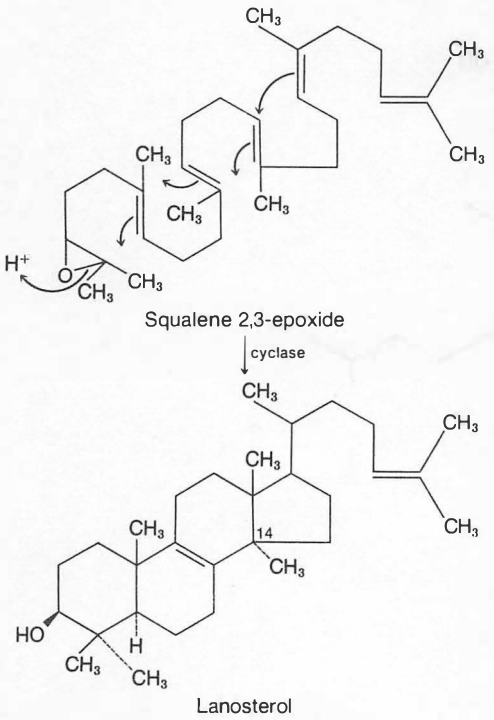


Figure 10.45
The conversion of squalene 2,3-epoxide to lanosterol.

undergoes two 1,2-methyl shifts, and a proton is eliminated. The oxygen atom is derived from molecular oxygen. The reaction is catalyzed by a microsomal enzyme system called squalene oxidocyclase that appears to be composed of at least two enzymes, an epoxidase or monoxygenase and a cyclase (lanosterol cyclase).

The cyclization process is initiated by epoxide formation between what will become C_2 and C_3 of cholesterol, the epoxide being formed at the expense of NADPH:

$$\text{Squalene} + O_2 + \text{NADPH} + H^+ \rightarrow \text{squalene 2,3-epoxide} + H_2O + NADP^+$$

This reaction is catalyzed by the monoxygenase or epoxidase component of squalene oxidocyclase. Hydroxylation at C-3 by way of the epoxide intermediate triggers the cyclization of squalene to form lanosterol as shown in Figure 10.45. In the process of cyclization, two hydrogen atoms and two methyl groups migrate to neighboring positions.

The transformation of lanosterol to cholesterol involves many steps and a number of different enzymes. These steps include (1) removal of the methyl group at C-14, (2) removal of the two methyl groups at C-4, (3) migration of the double bond from C-8 to C-5, and (4) reduction of the double bond between C-24 and C-25 in the side chain. (See Figure 10.46).

Regulation of Cholesterol Biosynthesis

The cholesterol pool of the body is derived from two sources: absorption of dietary cholesterol and biosynthesis de novo, primarily in the liver and the intestine. When the amount of dietary cholesterol is

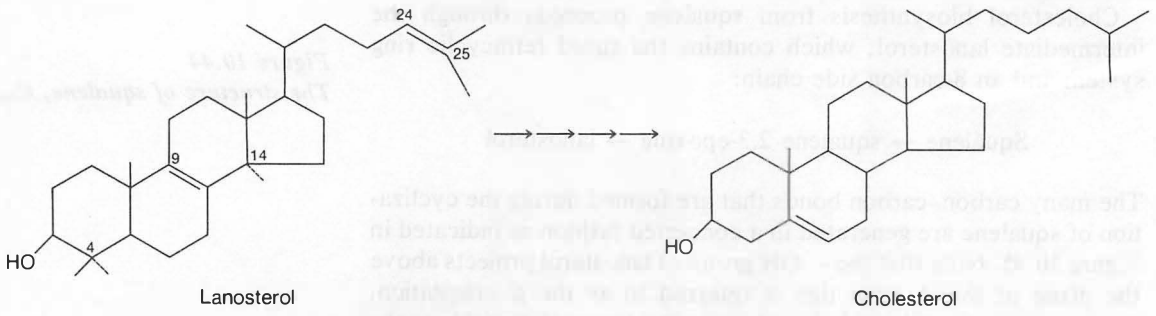


Figure 10.46
The conversion of lanosterol to cholesterol.

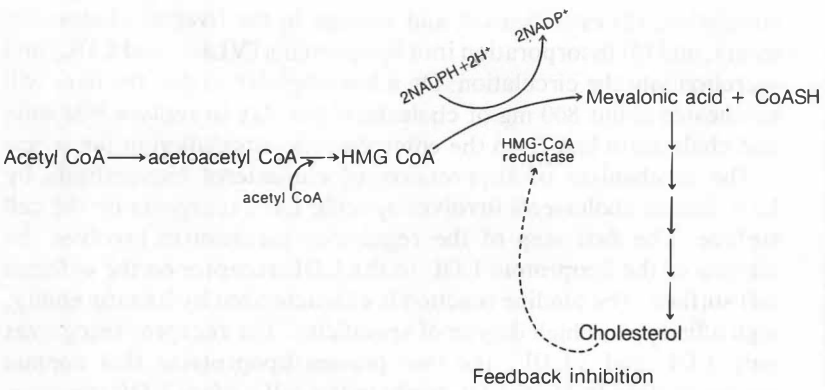


Figure 10.47
Summary of the pathway of cholesterol synthesis indicating feedback inhibition of HMG CoA reductase by cholesterol.

reduced, cholesterol synthesis is increased in the liver and intestine to satisfy the needs of other tissues and organs. Cholesterol synthesized de novo is transported from the liver and intestine to peripheral tissues in the form of lipoproteins. These two tissues are the only ones that can manufacture apolipoprotein B, the protein component of the cholesterol transport proteins LDL and VLDL. Most of the apolipoprotein B is secreted into the circulation as VLDL, which is converted into LDL by removal of triglyceride and the apoprotein C components, probably in peripheral tissues and the liver. In contrast, when the quantity of dietary cholesterol increases, cholesterol synthesis in the liver and intestine is almost totally suppressed.

The primary site for control of cholesterol biosynthesis is the cytoplasmic isoenzyme of HMG CoA reductase, which catalyzes the step that produces mevalonic acid. This is the committed step and the rate-limiting reaction in the pathway of cholesterol biosynthesis (Figure 10.47). Cholesterol effects feedback inhibition of its own synthesis by inhibiting the activity of preexisting HMG CoA reductase and also by promoting rapid inactivation of the enzyme by mechanisms that remain to be elucidated. The dietary cholesterol that suppresses HMG CoA reductase activity and cholesterol synthesis emerges from the intestine in the form of chylomicrons.

In a normal healthy adult on a low cholesterol diet about 1,300 mg of cholesterol is returned to the liver each day for disposition. This cholesterol comes from (1) cholesterol reabsorbed from the gut by means of the enterohepatic circulation and (2) HDL that carries cholesterol to the liver from peripheral tissues. The liver disposes of cholesterol in one of three ways: (1) excretion in bile as free cholesterol and after conversion to bile salts—each day, about 250 mg of bile salts and 550 mg of cholesterol are lost from the enterohepatic

CLIN. CORR. 10.2 ATHEROSCLEROSIS

Atherosclerosis is a complex and chronic disease involving the gradual accumulation of lipids, collagen, elastic fibers, and proteoglycans in the arterial wall. Since cholesterol esters and cholesterol are major components of atherosclerotic lesions, the interaction of the cholesterol-carrying lipoproteins in plasma with the cells of the arterial wall seem to be important. An increased level of total plasma cholesterol and an increase in the major cholesterol-carrying lipoprotein, LDL, are associated with an increased risk of developing atherosclerotic cardiovascular disease. There is also considerable evidence that an elevated level of plasma triglyceride is also a risk factor for ischemic heart disease. On the other hand, high concentrations of HDL seem to decrease the risk for ischemic heart disease.

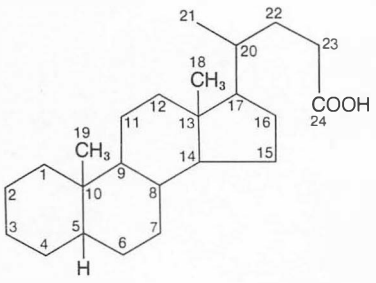


Figure 10.48
The structure of cholanic acid.

circulation, (2) esterification and storage in the liver as cholesterol esters, and (3) incorporation into lipoproteins (VLDL and LDL) and secretion into the circulation. On a low-cholesterol diet the liver will synthesize about 800 mg of cholesterol per day to replace bile salts and cholesterol lost from the enterohepatic circulation in the feces.

The mechanism of suppression of cholesterol biosynthesis by LDL-bound cholesterol involves specific LDL receptors on the cell surface. The first step of the regulatory mechanism involves the binding of the lipoprotein LDL to the LDL receptor on the external cell surface. The binding reaction is characterized by its saturability, high affinity, and high degree of specificity. The receptor recognizes only LDL and VLDL, the two plasma lipoproteins that contain apolipoprotein B. Once LDL binds to the cell surface LDL receptor, the cholesterol—charged lipoprotein is endocytosed in the form of vesicles called endosomes. The next step involves the fusion of the endosome with a lysosome that contains numerous hydrolytic enzymes, including proteases and cholesterol esterase. Inside the lysosome the cholesterol ester component of LDL is hydrolyzed by lysosomal cholesterol esterase to produce free cholesterol and a molecule of long-chain fatty acid. Free cholesterol then diffuses into the cytoplasm where it inhibits the activity of microsomal HMG CoA reductase by some unknown mechanism and surpresses the synthesis of HMG CoA reductase enzyme. There is evidence that cholesterol acts at the level of DNA and the protein synthesis apparatus to decrease the rate of synthesis of HMG CoA reductase protein. At the same time, microsomal fatty acyl CoA : cholesterol acyltransferase (ACAT) is activated by cholesterol, thereby promoting the formation of cholesterol esters, principally cholesterol oleate. The accumulation of intracellular cholesterol esters eventually inhibits the replenishment of LDL receptors on the cell surface, a phenomenon called "down-regulation," thereby blocking further uptake and accumulation of cholesterol by the cell.

Patients with familial (genetic) hypercholesterolemia suffer from accelerated arteriosclerosis (Clin. Corr. 10.2) and have a defect in this regulatory system. In such individuals, referred to as receptor-negative, cells lack functional LDL receptors on the surfaces. As a result, there is no binding of LDL to the cell, cholesterol is not transferred into the cell, and cholesterol synthesis is not inhibited.

Bile Acids and the Excretion of Cholesterol

The bile acids are the end products of cholesterol metabolism. Primary bile acids are those that are synthesized in hepatocytes directly from cholesterol. The most abundant bile acids in human beings are derivatives of cholanic acid (Figure 10.48). The most common bile

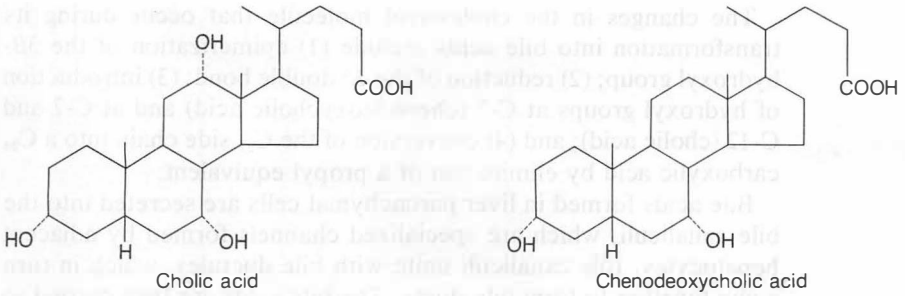


Figure 10.49
The structures of two common bile acids.

acids in human beings are cholic acid and chenodeoxycholic acid (Figure 10.49).

One should keep in mind that the primary bile acids (1) are composed of 24 carbon atoms, (2) are made in liver parenchymal cells, (3) contain two or three hydroxyl groups, and (4) have a side chain that ends in a carboxyl group that is ionized at pH 7.0 (hence the name bile salt). The carboxyl group of the primary bile acids is often conjugated via an amide bond to either glycine (NH_2—CH—COOH) or taurine (NH_2—CH_2—CH_2—SO_3H) to form glycocholic or taurocholic acid, respectively. The structure of glycocholic acid is shown in Figure 10.50.

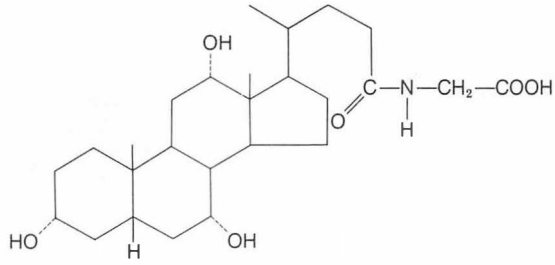


Figure 10.50
The structure of glycocholic acid.

When the primary bile acids undergo further chemical reactions by microorganisms in the gut, they give rise to secondary bile acids that also possess 24 carbon atoms. Examples of secondary bile acids are deoxycholic acid and lithocholic acid, which are derived from cholic acid and chenodeoxycholic acid, respectively, by the removal of one hydroxyl group (Figure 10.51).

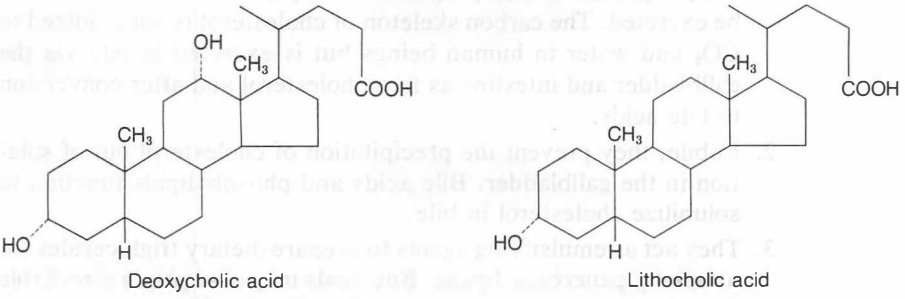


Figure 10.51
The structures of two secondary bile acids.

The changes in the cholesterol molecule that occur during its transformation into bile acids include (1) epimerization of the 3β-hydroxyl group; (2) reduction of the Δ^5 double bond; (3) introduction of hydroxyl groups at C-7 (chenodeoxycholic acid) and at C-7 and C-12 (cholic acid); and (4) conversion of the C_{27} side chain into a C_{24} carboxylic acid by elimination of a propyl equivalent.

Bile acids formed in liver parenchymal cells are secreted into the bile canaliculi, which are specialized channels formed by adjacent hepatocytes. Bile canaliculi unite with bile ductules, which in turn come together to form bile ducts. The bile acids are then carried to the gallbladder for storage and ultimately to the small intestine where they are excreted. The capacity of the liver to produce bile acids is insufficient to meet the physiological demands, so the body relies upon an efficient enterohepatic circulation that carries the bile acids from the intestine back to the liver several times each day. The primary bile acids, after removal of the glycine or taurine residue in the gut, are reabsorbed by an active transport process from the intestine, primarily in the ileum, and returned to the liver by way of the portal vein. Bile acids that are not reabsorbed are acted upon by bacteria in the gut and converted into secondary bile acids: a portion of secondary bile acids, primarily deoxycholic acid and lithocholic acid, are reabsorbed passively in the colon and returned to the liver where they are secreted into the gallbladder. Hepatic synthesis normally produces 0.2–0.6 g of bile acids per day to replace those lost in the feces. The gallbladder pool of bile acids is 2–4 g. Because the enterohepatic circulation recycles 6 to 12 times each day, the total amount of bile acids absorbed per day from the intestine corresponds to 12–32 g.

In terms of function the bile acids are significant in medicine for several reasons:

1. They represent the only significant way in which cholesterol can be excreted. The carbon skeleton of cholesterol is not oxidized to CO_2 and water in human beings but is excreted in bile via the gallbladder and intestine as free cholesterol and after conversion to bile acids.
2. In bile, they prevent the precipitation of cholesterol out of solution in the gallbladder. Bile acids and phospholipids function to solubilize cholesterol in bile.
3. They act as emulsifying agents to prepare dietary triglycerides for attack by pancreatic lipase. Bile acids may also play a direct role in activating pancreatic lipase. (See Chapter 24.)
4. They facilitate the absorption of fat-soluble vitamins, particularly vitamin D, from the intestine.

10.4 SPHINGOLIPIDS

Structure and Biosynthesis of Sphingosine and Ceramide

Sphingolipids are complex lipids whose core structure is provided by the long-chain aminoalcohol sphingosine (Figure 10.52). Another common name for sphingosine is 4-sphingenine; the formal name for sphingosine is *trans*-1,3-dihydroxy-2-amino-4-octadecene. We shall learn that sphingosine (1) possesses two asymmetric carbon atoms (positions C-2 and C-3); of the four possible optical isomers, naturally occurring sphingosine is of the D-erythro form; (2) the Δ^4 carbon–carbon double bond has the trans configuration; (3) the primary alcohol group at C-1 is a nucleophilic center that forms covalent bonds with sugars to form glycosides called glycolipids or combines with phosphocholine to form sphingomyelin; (4) the amino group at C-2 always bears a long-chain (usually C_{20}–C_{26}) fatty acid in amide linkage; and (5) the secondary alcohol at C-3 is never derivatized and is always free. It is useful to appreciate the structural similarity of a part of the sphingosine molecule to the glycerol moiety of the acyl glycerides. When one views the structure of sphingosine from another perspective, the similarity between carbons 1, 2, and 3 of sphingosine and glycerol becomes apparent (Figure 10.53). Note that both glycerol and sphingosine have nucleophilic groups, hydroxyl or amino, at positions C-1, C-2, and C-3.

The sphingolipids occur in blood and nearly all of the tissues of human beings. However, the highest concentrations of sphingolipids are found in the white matter of the central nervous system. Various sphingolipids are components of the plasma membrane of practically all cells.

Sphingosine is synthesized by way of sphinganine (dihydrosphingosine) in two steps from the precursors L-serine and palmitoyl CoA: serine is the source of C-1, C-2, and the amino group of sphingosine, while palmitic acid provides the remaining carbon atoms. The condensation of serine with palmitoyl CoA is catalyzed by a pyridoxal phosphate-dependent enzyme and the driving force for the reaction is provided by both the cleavage of the reactive, high energy, thioester bond of palmitoyl CoA and the release of CO_2 from serine (Figure 10.54). The next step involves the reduction of the carbonyl group in 3-keto dihydrosphingosine with reducing equivalents being derived from NADPH (Figure 10.55). The insertion of the double bond into sphinganine to produce sphingosine occurs at the level of ceramide (see below).

Sphingosine as such with its free amino group does not occur

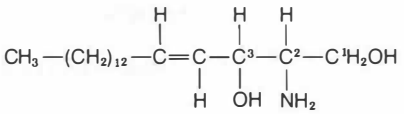


Figure 10.52
***The structure of sphingosine* (trans-*1,3-dihydroxy-2-amino-4-octadecene*).**

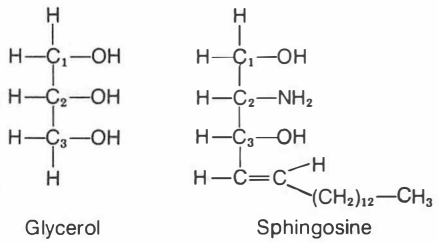


Figure 10.53
Comparison of the structures of glycerol and sphingosine.

fig. 10.54

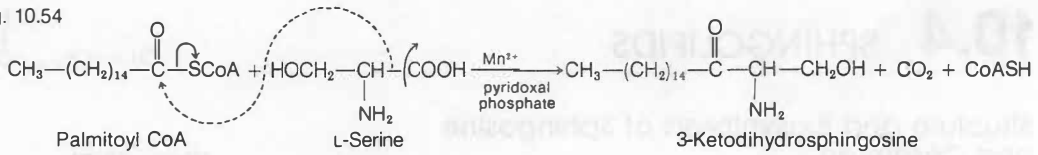


Figure 10.54
The formation of 3-ketodihydrosphingosine from serine and palmitoyl CoA.

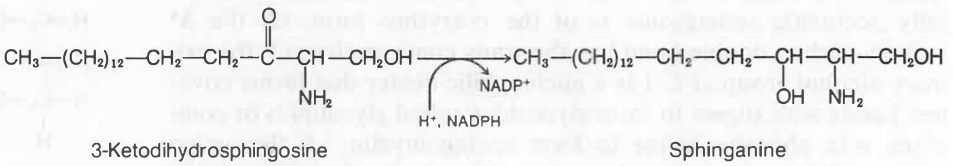


Figure 10.55
The conversion of 3-ketodihydrosphingosine to sphinganine.

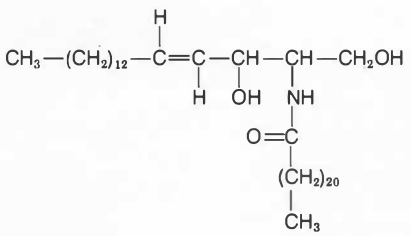


Figure 10.56
The structure of ceramide.

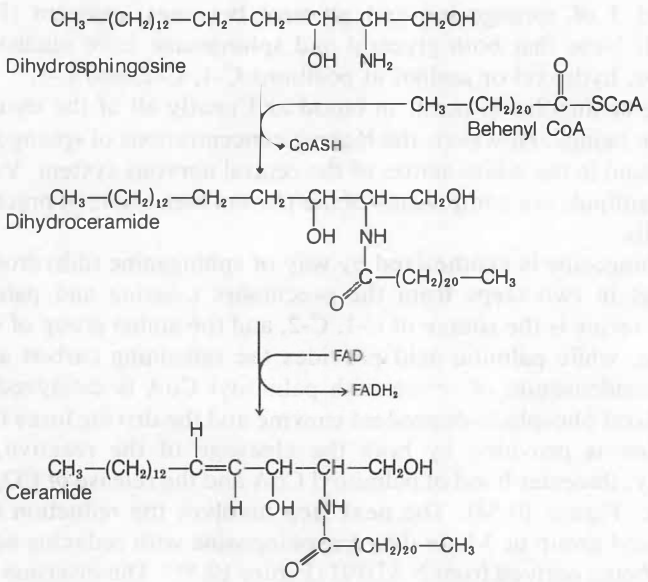
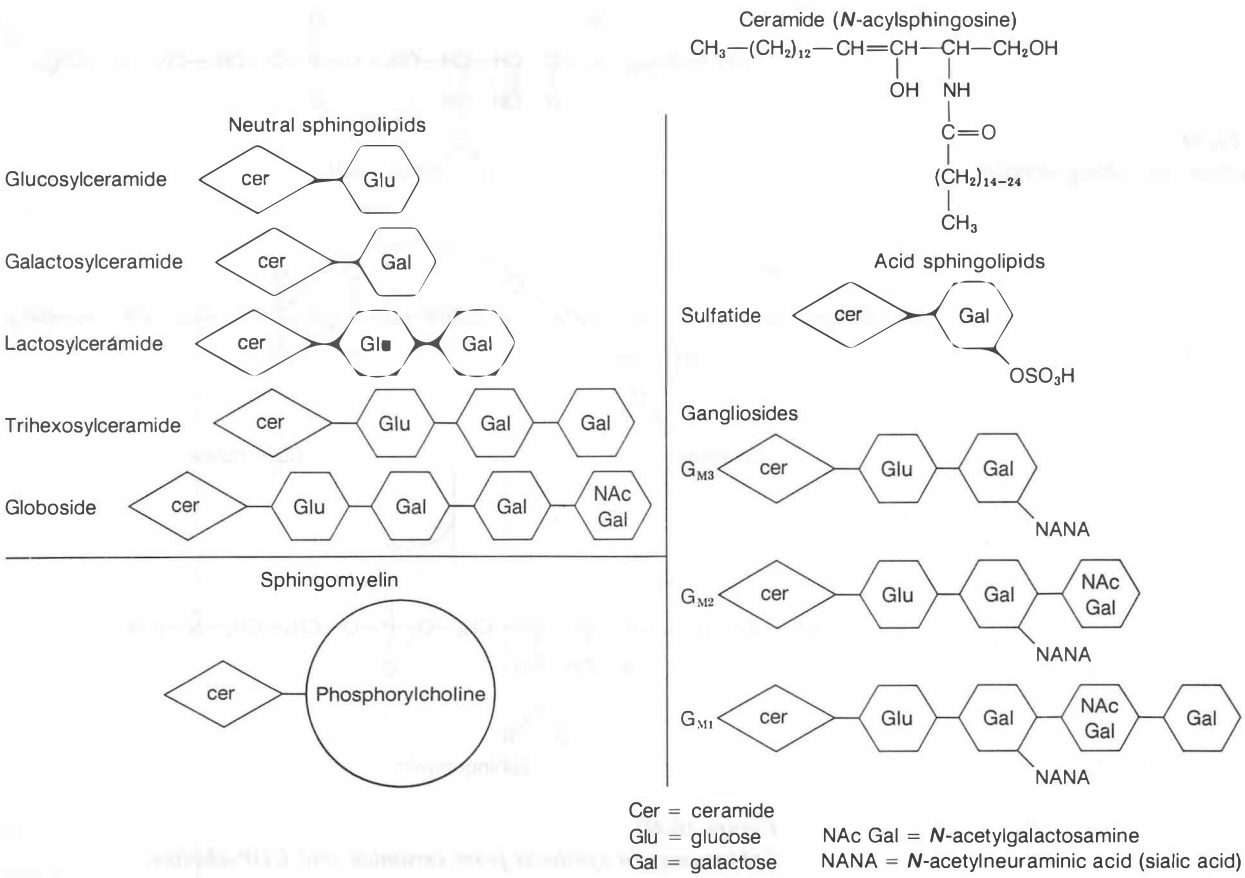


Figure 10.57
The formation of ceramide from dihydrosphingosine.

naturally. The fundamental building block or core structure of the natural sphingolipids is ceramide. Ceramide, also called cerebroside, is the long-chain fatty acid amide derivative of sphingosine. Most often the acyl group is behenic acid, a saturated C_{22} fatty acid. Note that the long-chain fatty acid is attached to the 2-amino group of sphingosine through an amide bond (Figure 10.56). Also note the presence of two long-chain hydrocarbon domains in the ceramide molecule; these hydrophobic regions are responsible for the lipoidal character of the sphingolipids.

Ceramide is synthesized from dihydrosphingosine and a molecule of long-chain fatty acyl CoA in a reaction catalyzed by a microsomal enzyme with the intermediary of dihydroceramide (Figure 10.57). Free ceramide is not a component of membrane lipids but rather represents an intermediate in the biosynthesis and catabolism of glycosphingolipids and sphingomyelin. Figure 10.58 contains, in dia-

Figure 10.58
The structures of some of the common sphingolipids in diagrammatic form.



grammatic form, the structures of the prominent sphingolipids of humans.

Sphingomyelin

Sphingomyelin is one of the principal structural lipids of the membranes of nervous tissue. It is the only sphingolipid that is a phospholipid, and it is not a glycolipid. In sphingomyelin the primary alcohol group at C-1 of sphingosine is esterified to choline through a phosphodiester bridge of the kind that occurs in the acyl glycerophospholipids and the amino group of sphingosine is attached to a long-chain fatty acid by means of an amide bond. Sphingomyelin is therefore a ceramide-phosphocholine. Note that sphingomyelin contains one negative and one positive charge so that it represents a neutral compound at physiological pH (Figure 10.59).

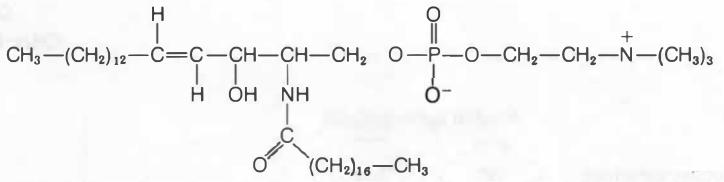


Figure 10.59
The structure of sphingomyelin.

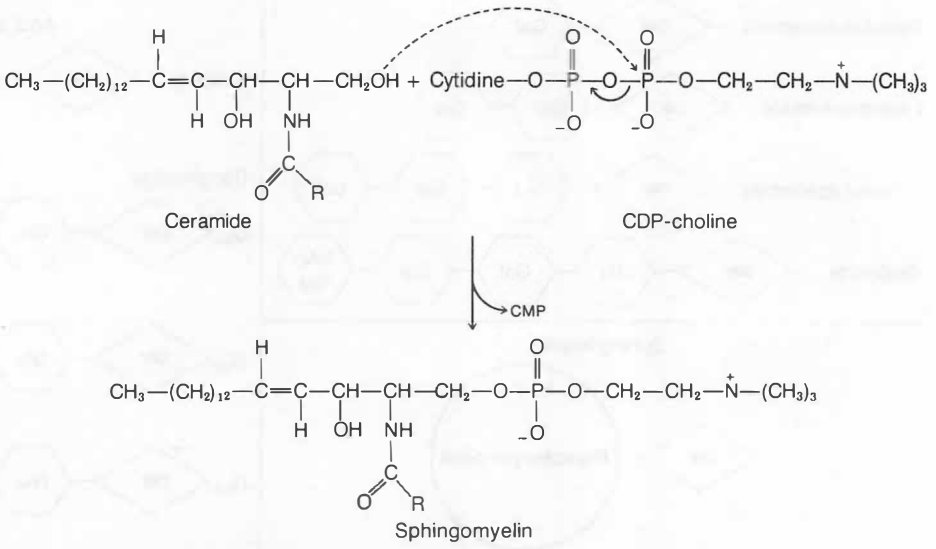


Figure 10.60
Sphingomyelin synthesis from ceramide and CDP-choline.

The most common fatty acids in sphingomyelin are palmitic, stearic, lignoceric (a C_{24}, saturated fatty acid), and nervonic acid [24: 1, $CH_3—(CH_2)_7CH=CH—(CH_2)_{13}—COOH$]. The sphingomyelin of myelin contains predominantly longer chain fatty acids, mainly lignoceric and nervonic, whereas that of grey matter contains largely stearic acid. Excessive accumulations of sphingomyelin occur in Niemann–Pick disease.

Sphingomyelin Synthesis

The conversion of ceramide to sphingomyelin involves the transfer of the phosphocholine group from CDP-choline to the primary, C-1 hydroxyl group of ceramide: this reaction is catalyzed by the enzyme CDP-choline : ceramide cholinephosphotransferase (Figure 10.60).

Carbohydrate-containing Sphingolipids

The principal glycosphingolipid classes are cerebrosides, sulfatides, globosides, and gangliosides. In the glycolipid class of compounds the polar head group is attached to sphingosine via the glycosidic linkage of a sugar molecule rather than a phosphate ester bond, as is the case in the phospholipids.

Cerebrosides

The cerebrosides are a group of ceramide monohexosides. The two most common cerebrosides of 1-β-glycosylceramides encountered in medicine are galactocerebroside and glucocerebroside. Unless specified otherwise, the term cerebroside usually refers to galactocerebroside. Galactocerebroside is also called "galactolipid." In Figure 10.61 note that the monosaccharide units are attached at C-1 of the sugar moiety to the C-1 position of ceramide, and the anomeric configuration of the glycosidic bond between ceramide and hexose in both galactocerebroside and glucocerebroside is β (beta). The largest amount of galactocerebroside in healthy people is found in the brain. Moderately increased amounts of galactocerebroside accumulate in the white matter in Krabbe's disease, also called globoid leukodystrophy, due to a deficiency in the lysosomal enzyme galactocerebrosidase.

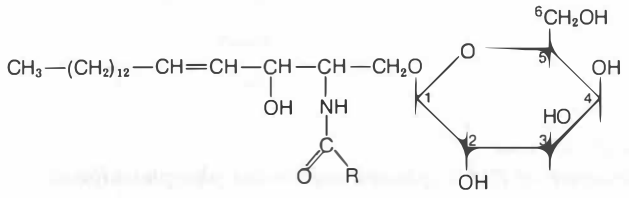


Figure 10.61
The structure of galactocerebroside (galactolipid).

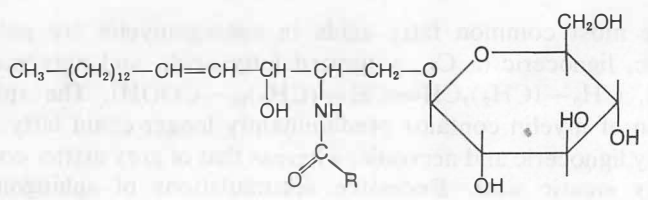


Figure 10.62
The structure of glucocerebroside.

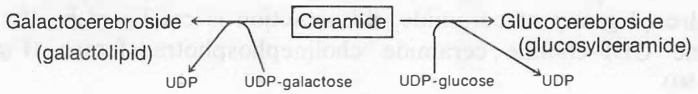


Figure 10.63
The synthesis of galacto- and glucocerebroside.

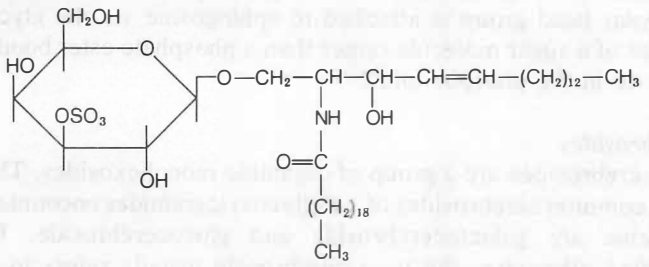


Figure 10.64
The structure of galactocerebroside sulfate (sulfolipid).

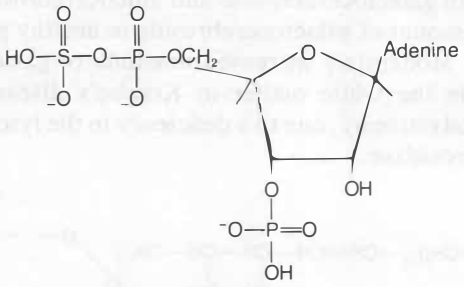


Figure 10.65
The structure of PAPS (phosphoadenosine phosphosulfate).

Glucocerebroside is not normally a structural component of membranes and is observed only during the synthesis and degradation of more complex glycosphingolipids. However, hundredfold increases in the glucocerebroside content of spleen and liver occur in the genetic lipid storage disorder called Gaucher's disease, which results from a deficiency of lysosomal glucocerebrosidase (See Figure 10.62).

Galactocerebroside and glucocerebroside are synthesized from ceramide and the activated sugar nucleotides UDP-galactose and UDP-glucose, respectively. The enzymes that catalyze these reactions, glucosyl and galactosyl transferases, are associated with the endoplasmic reticulum (Figure 10.63). Alternatively, in some tissues, the synthesis of glucocerebroside (glucosylceramide) proceeds by way of glucosylation of sphingosine:

$$\text{Sphingosine} + \text{UDP-glucose} \xrightarrow[\text{glucosyltransferase}]{} \text{glucosylsphingosine} + \text{UDP}$$

followed by fatty acylation:

$$\text{Glucosylsphingosine} + \text{stearoyl CoA} \dashrightarrow \text{glucocerebroside} + \text{CoASH}$$

Sulfatide

Sulfatide, or sulfocerebroside as it is sometimes called, is a sulfuric acid ester of galactocerebroside. Galactocerebroside 3-sulfate is the major sulfolipid in brain and accounts for approximately 15% of the lipids of white matter. (See Figure 10.64).

Galactocerebroside sulfate is synthesized from galactocerebroside and "activated sulfate" or PAPS (3′-phosphoadenosine 5′-phosphosulfate) in a reaction catalyzed by microsomal sulfotransferase:

$$\text{Galactocerebroside} + \text{PAPS} \rightarrow \text{PAP} + \text{galactocerebroside 3-sulfate}$$

The structure of PAPS is indicated in Figure 10.65. Large quantities of sulfatide accumulate in the tissues of the central nervous system in metachromatic leukodystrophy due to a deficiency in a sulfatase enzyme.

Globosides—Ceramide Oligosaccharides

This family of compounds represents cerebrosides that contain two or more sugar residues, usually galactose, glucose, or *N*-acetylgalactosamine. The ceramide oligosaccharides are neutral compounds and contain no free amino groups.

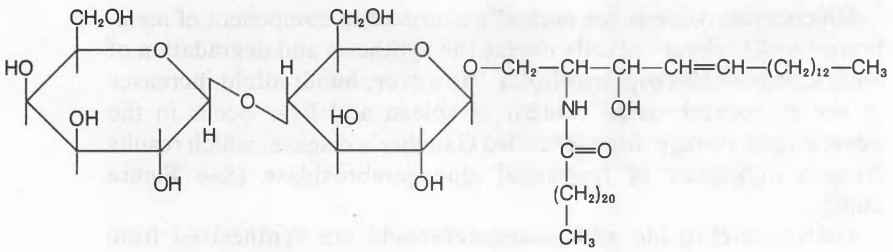


Figure 10.66
The structure of ceramide-β-glc-(4 ⟵ 1)-β-gal-(4 ⟵ 1) (lactosylceramide).

Lactosylceramide is a component of the erythrocyte membrane (Figure 10.66).

Another prominent compound in this series of glycosphingolipids is ceramide trihexoside or ceramide galactosyllactoside: ceramide-β-glc-(4 ← 1)-β-gal-(4 ← 1)-α-gal. Note that the terminal galactose residue of this globoside has the α-anomeric configuration. Ceramide trihexoside accumulates in the kidneys of patients with Fabry's disease who are deficient in lysosomal α-galactosidase activity.

Gangliosides

The name ganglioside was adopted for the class of sialic acid-containing glycosphingolipids that are highly concentrated in the ganglion cells of the central nervous system, particularly in the nerve endings. The central nervous system is unique among human tissues because more than half of the sialic acid is in ceramide lipid-bound form, with the remainder of the sialic acid occurring in the oligosaccharides of glycoproteins. Lesser amounts of gangliosides are contained in the surface membranes of the cells of most extraneural tissues where they account for less than 10% of the total sialic acid.

Neuraminic acid (abbreviated Neu) is present in gangliosides, glycoproteins, and mucins. The amino group of neuraminic acid occurs most often as the *N*-acetyl derivative, and the resulting structure is called *N*-acetylneuraminic acid or sialic acid, commonly abbreviated NANA. (See Figure 10.67).

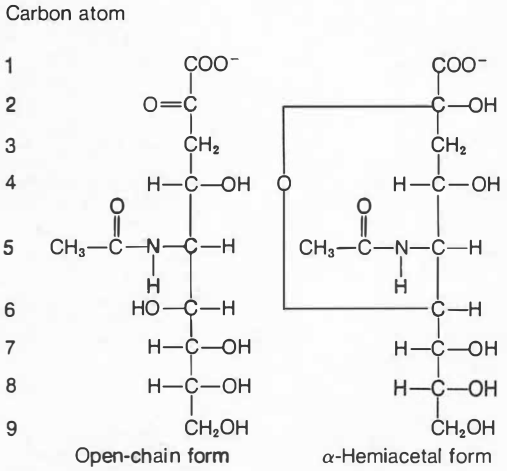


Figure 10.67
The structure of **N-acetylneuraminic acid** *(NANA).*

The hydroxyl group on C-2 occurs most often in the α-anomeric configuration and the linkage between NANA and the oligosaccharide ceramide always involves the —OH group on position-2 of *N*-acetylneuraminic acid.

The structures of some of the common gangliosides is indicated in Table 10.1. The principal gangliosides in brain are G_{M1}, G_{D1a}, G_{D1b}, and G_{T1}. Nearly all of the gangliosides of man are derived from the family of compounds originating with glucosylceramide.

With regard to the nomenclature of the sialoglycosylsphingolipids, the letter G refers to the name ganglioside. The subscripts M, D, T,

Table 10.1 The Structures of Some Common Gangliosides

Code Name	*Chemical Structure*
G_{M3}	Galβ → 4Glcβ → Cer 3 ↑ αNANA
G_{M2}	GalNAcβ → 4Galβ → 4Glcβ → Cer 3 ↑ αNANA
G_{M1}	Galβ → 3GalNAcβ → 4Galβ → 4Glcβ → Cer 3 ↑ αNANA
G_{D1a}	Galβ → 3GalNAcβ → 4Galβ → 4Glcβ → Cer 3 3 ↑ ↑ αNANA αNANA
G_{D1b}	Galβ → 3GalNAcβ → 4Galβ → 4Glcβ → Cer 3 ↑ αNANA8 ← αNANA
G_{T1}	Galβ → 3GalNAcβ → 4Galβ → 4Glcβ → Cer 3 3 ↑ ↑ αNANA αNANA8 ← αNANA

and Q indicate mono-, di-, tri-, and quatra(tetra)-sialic acid-containing gangliosides. The numerical subscripts 1, 2, and 3 designate the carbohydrate sequence that is attached to ceramide as indicated as follows: 1, Gal-GalNAc-Gal-Glc-ceramide; 2, GalNAc-Gal-Glc-ceramide; and 3, Gal-Glc-ceramide. Consider the nomenclature of the Tay–Sachs ganglioside; the designation G_{M2} denotes the ganglioside structure shown in Figure 10.68.

A specific ganglioside on intestinal mucosal cells mediates the action of cholera toxin. Cholera toxin is a protein of mol w 84,000 that is secreted by the pathogen *Vibrio cholerae*. The toxin stimulates the secretion of chloride ions into the gut lumen, resulting in the severe diarrhea characteristic of cholera. Two kinds of subunits, A and B, comprise the cholera toxin; there is one copy of the A subunit (28,000 daltons) and six copies of the B subunit (~12,000 daltons each). After binding to the cell surface membrane through a domain on the B subunit, the active subunit A passes into the cell, where it activates adenylate cyclase on the inner surface of the membrane.

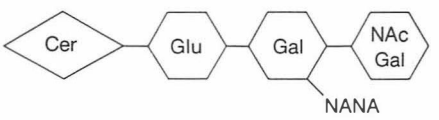


Figure 10.68
The structure of the Tay–Sachs ganglioside, G_{M2}.

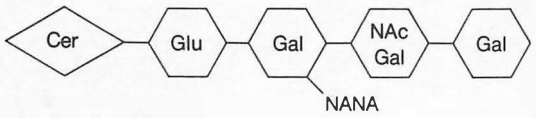


Figure 10.69
The structure of G_{M1}.

The cAMP that is generated then stimulates chloride ion transport and produces diarrhea. The choleragenoid domain, as the B subunits are called, binds to the ganglioside G_{M1} that has the structure shown in Figure 10.69.

Gangliosides are also thought to be receptors for other toxic agents, such as tetanus toxin, and certain viruses, such as the influenza viruses. There is also speculation that gangliosides may play an informational role in cell–cell interactions by providing specific recognition determinants on the surface of cells.

The gangliosides are also of medical interest for there are several lipid storage disorders that involve the accumulation of sialic acid-containing glycosphingolipids. The two most common gangliosidoses involve the storage of the gangliosides G_{M1} (G_{M1} gangliosidosis) and G_{M2} (Tay–Sachs disease).

G_{M1} gangliosidosis is an autosomal recessive metabolic disease characterized by impaired psychomotor function, mental retardation, hepatosplenomegaly, and death within the first few years of life. The massive cerebral and visceral accumulation of G_{M1} ganglioside is due to a profound deficiency of β-galactosidase.

The Sphingolipidoses–Lysosomal Storage Disease and the Catabolic Pathway of the Sphingolipids

The various sphingolipids are normally degraded within lysosomes of phagocytic cells, particularly the histiocytes or macrophages of the reticuloendothelial system located primarily in the liver, spleen, and bone marrow. Degradation of the sphingolipids by visceral organs begins with the engulfment of the membranes of white cells and erythrocytes that are rich in lactosylceramide (Cer-Glc-Gal) and hematoside (Cer-Glc-Gal-NANA). In the brain, the majority of the cerebroside-type lipids are gangliosides. Particularly during the neonatal period, ganglioside turnover in the central nervous system is extensive so that glycosphingolipids are rapidly being broken down and resynthesized. The pathway of sphingolipid catabolism is summarized in Figure 10.70. Note that among the various sphingolipids that comprise this pathway, there occurs a sulfate ester (in sulfolipid or sulfogalactolipid); *N*-acetylneuraminic acid groups (in the gangliosides); an α-linked galactose residue (in ceramide trihexoside); several β-galactosides (in galactocerebroside and the ganglioside G_{M1}); the ganglioside G_{M2}, which terminates in a β-linked *N*-acetylgalactosamine unit; and glucocerebroside, which is composed of a single glucose residue attached to ceramide through a β linkage. We also see that the phosphodiester bond in sphingomyelin is broken to produce ceramide, which in turn is converted to sphin-

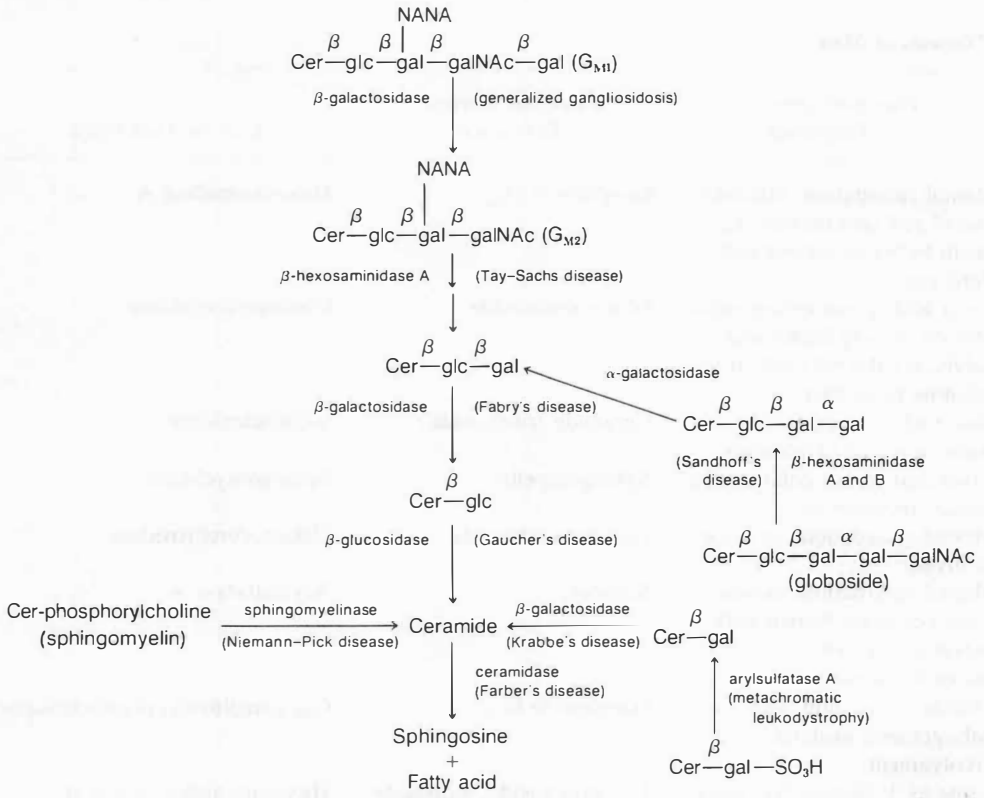


Figure 10.70
Summary of the pathways for the catabolism of sphingolipids by lysosomal enzymes.
The genetically determined enzyme deficiency diseases are indicated in the parentheses.

gosine by the cleavage of an amide bond to a long-chain fatty acid. This overall pathway of sphingolipid catabolism is composed of a series of enzymes that cleave specific bonds in the compounds that comprise the pathway: these enzymes include α- and β-galactosidases, a β-glucosidase, a neuraminidase (also called sialidase), hexosaminidase, a sphingomyelin-specific phosphodiesterase (sphingomyelinase), a sulfate esterase (sulfatase), and a ceramide-specific amidase (ceramidase). The following statements summarize the important features of the sphingolipid catabolic pathway: (1) the reactions all take place within the lysosome; that is, the enzymes of the pathway are contained in lysosomes; (2) the enzymes are hydrolases; therefore, one of the substrates in each reaction is water; (3) the pH optimum of each of the hydrolases is in

Table 10.2 Sphingolipid Storage Diseases of Man

Disorder	*Principal Signs and Symptoms*	*Principal Storage Substance*	*Enzyme Deficiency*
1. Tay–Sachs disease	Mental retardation, blindness, cherry red spot on macula, death between second and third year	Ganglioside G_{M2}	Hexosaminidase A
2. Gaucher's disease	Liver and spleen enlargement, erosion of long bones and pelvis, mental retardation in infantile form only	Glucocerebroside	Glucocerebrosidase
3. Fabry's disease	Skin rash, kidney failure, pains in lower extremities	Ceramide trihexoside	α-Galactosidase
4. Niemann–Pick disease	Liver and spleen enlargement, mental retardation	Sphingomyelin	Sphingomyelinase
5. Globoid leukodystrophy (Krabbe disease)	Mental retardation, absence of myelin	Galactocerebroside	Galactocerebrosidase
6. Metachromatic leukodystrophy	Mental retardation, nerves stain yellowish brown with cresyl violet dye (metachromasia)	Sulfatide	Arylsulfatase A
7. Generalized gangliosidoses	Mental retardation, liver enlargement, skeletal involvement	Ganglioside G_{M1}	G_{M1} ganglioside, β-galactosidase
8. Sandhoff–Jatzkewitz disease	Same as 1; disease has more rapidly progressing course	G_{M2} ganglioside, globoside	Hexosaminidase A and B
9. Fucosidosis	Cerebral degeneration, muscle spasticity, thick skin	Pentahexosylfucoglycolipid	α-L-Fucosidase

the acidic range, pH 3.5–5.5; (4) most of the enzymes are relatively stable and occur as isoenzymes. For example, hexosaminidase occurs in two forms: hexosaminidase A (HexA) and hexosaminidase B (HexB); (5) the hydrolases of the sphingolipid pathway are glycoprotein in character and often occur firmly bound to the lysosomal membrane; and (6) the pathway is composed of a series of compounds that are related to the nearest compound in the pathway and which differ by only one sugar molecule, a sulfate group, or a fatty acid residue. That is, the substrates are converted to products by the sequential, stepwise removal of constituents such as sugars and sulfate, in hydrolytic, irreversible reactions.

In most cases, the pathway of sphingolipid catabolism functions smoothly, and all of the various complex glycosphingolipids and sphingomyelin are degraded to the level of their basic building blocks, namely, sugars, sulfate, fatty acid, phosphocholine, and

sphingosine. However, when the activity of one of the hydrolytic enzymes in the pathway is markedly reduced in the tissues of a person due to a genetic, inborn error, then the substrate for the defective or missing enzyme accumulates and deposits within the lysosomes of the tissue responsible for the catabolism of that lipid. For most of the reactions in Figure 10.70, patients have been identified who lack the enzyme that normally catalyzes that reaction. These disorders are called sphingolipidoses. Table 10.2 summarizes the individual diseases that comprise the sphingolipidoses.

One can generalize about some of the common features of lipid storage diseases: (1) usually only a single sphingolipid accumulates in the involved organs; (2) the ceramide portion is shared by the various storage lipids; (3) the rate of biosynthesis of the accumulating lipid is normal; (4) a catabolic enzyme is missing in each of these disorders; and (5) the extent of the enzyme deficiency is the same in all tissues.

The diagnosis of a given sphingolipidosis can be made from a biopsy of the involved organ, usually bone marrow, liver, or brain, or on morphologic grounds on the basis of the highly characteristic appearance of the storage lipid within lysosomes. Biochemical methods involving enzyme assays are also widely used to confirm the diagnosis of a particular lipid storage disease. Of great practical value is the fact that, for most of the diseases, peripheral leukocytes and cultured skin fibroblasts express the relevant enzyme deficiency and can be used as a source of enzyme for diagnostic purposes. In some cases (e.g., Tay–Sachs disease) serum, and even tears, have been used as a source of enzyme for the diagnosis of a lipid storage disorder. Because the sphingolipid storage diseases for the most part are recessive in terms of their hereditary mode of transmission, and disease occurs only in homozygotes with a defect in both chromosomes, enzyme assays can also be used to identify carriers or heterozygotes. Let us consider some representative examples of the use of enzyme assays for diagnostic purposes.

In Niemann–Pick disease, the deficient enzyme is sphingomyelinase, which normally catalyzes the reaction shown in Figure 10.71. Sphingomyelin, radiolabeled in the methyl groups with carbon-14, provides a useful substrate for determining sphingomyelinase activity. Extracts of white blood cells from healthy, appropriate controls, will hydrolyze the labeled substrate and produce the water-soluble product, [^{14}C]phosphocholine. Extraction of the final incubation medium with an organic solvent such as chloroform will result in radioactivity in the upper, aqueous phase; the unused, lipidlike substrate sphingomyelin will be found in the chloroform phase. On the other hand, if the white blood cells were derived from a patient with Niemann–Pick disease, then after incu-

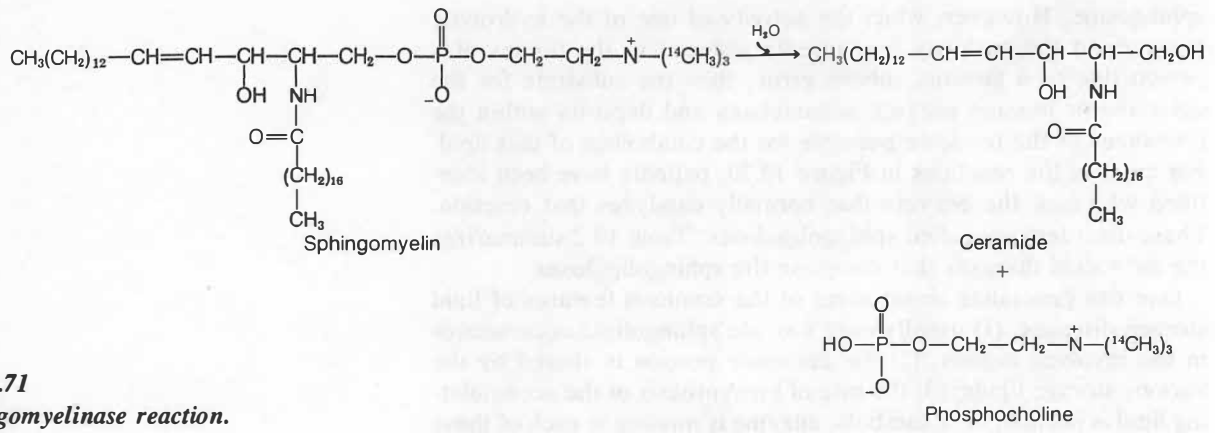


Figure 10.71
The sphingomyelinase reaction.

bation with labeled substrate and subsequent extraction with chloroform, little or no radioactivity (i.e., phosphocholine) will be found in the aqueous phase and the diagnosis of sphingomyelinase deficiency or Niemann–Pick disease will have been confirmed.

Because many hospitals or medical centers do not have access to radioactivity counting devices, it is fortunate that many of the lysosomal hydrolases have broad or versatile substrate specificity in that they will hydrolyze not only the true, natural sphingolipid substrate but also colorimetric or fluorometric nonphysiologic, unnatural synthetic substrates as well. For example, the sulfatidase deficiency in metachromatic leukodystrophy can be demonstrated, using the artificial substrate, nitrocatechol sulfate. The assay, called the aryl sulfatase determination, is indicated in Figure 10.72. In alkaline medium, *p*-nitrocatechol is red. Thus the amount of red color generated by incubating the enzyme extract with nitrocatechol sulfate can be quantitated with the aid of a spectrophotometer and is proportional to the level of sulfatase activity. The utility of the enzyme assay is due to the fact that sulfolipid sulfatase will act on the artificial, colorimetric substrate *p*-nitrocatechol sulfate.

Another sphingolipid storage disease that can be diagnosed using an artificial substrate is Tay–Sachs disease. Tay–Sachs disease is the most common form of G_{M2} gangliosidosis. In this fatal disorder the ganglion cells of the cerebral cortex are swollen and the lysosomes are engorged with the acidic lipid, G_{M2} ganglioside. This results in a loss of ganglion cells, proliferations of glial cells and demyelination of peripheral nerves. The pathopneumonic finding is a cherry-red spot on the macula caused by swelling and necrosis of ganglion cells in the eye. In Tay–Sachs disease, the commercially available artificial substrate 4-methylumbelliferyl-β-*N*-acetylglucosaminide is used to confirm the diagnosis on biochemical grounds. The compound is

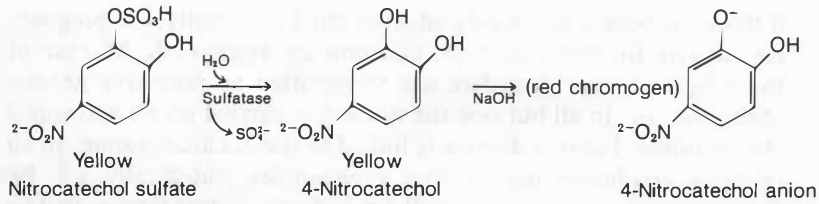
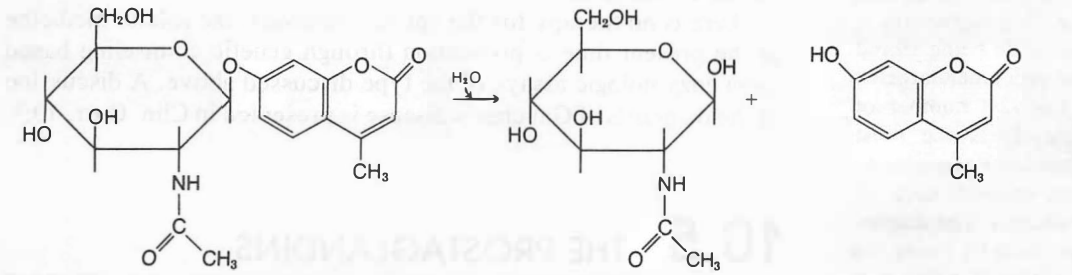


Figure 10.72
The arylsulfatase assay.



4-Methylumbelliferyl-β-D-*N*-acetylglucosaminide

N-Acetylglucosamine

4-Methylumbelliferone
(fluorescent in alkaline medium)

Figure 10.73
The β-hexosaminidase reaction.

recognized and hydrolyzed by hexosaminidase A, the deficient lysosomal hydrolase in Tay–Sachs disease, to produce the intensely fluorescent product 4-methylumbelliferone (Figure 10.73). Unfortunately, the diagnosis of Tay–Sachs disease may be confused by the presence of hexosaminidase B present in tissue extracts and body fluids that are used as a source of enzyme. Hexosaminidase B is not deficient in the Tay–Sachs patient and will hydrolyze the test substrate, thereby confusing the interpretation of results. The problem of isoenzymes is usually resolved by taking advantage of the relative heat lability of hexosaminidase A and the heat stability of hexosaminidase B. The tissue extract or serum specimen to be tested is first heated at 55°C for 1 h and then assayed for hexosaminidase activity. The amount of heat-labile activity is thus a measure of hexosaminidase A, and this value is used to consider the diagnosis of Tay–Sachs disease.

In addition to confirming the diagnosis of a particular lipid storage disease, enzyme assays of serum or extracts of tissues, peripheral leukocytes and fibroblasts have proven useful in heterozygote detection. Once carriers of a lipid storage disease have been identified, or

CLIN. CORR. 10.3 THE DIAGNOSIS OF GAUCHER'S DISEASE IN AN ADULT

Gaucher's disease is an inherited disease of cell membrane catabolism that results in deposits of glucocerebroside being stored in macrophages of the reticuloendothelial system. Because of the vast number of storage cells splenomegaly is the most common sign, and often some sequelae to hypersplenism will be evident such as thrombocytopenia or anemia. The diagnosis has most often been made by seeing the highly characteristic Gaucher cells in a sample of aspirated bone marrow.

A 61-year-old nurse (H.V.) of Ashkenazi Jewish ancestry complaining of upper abdominal discomfort whose spleen was barely palpable and who lacked evidence of hypersplenism was suspected of having Gaucher's disease. Because she refused to permit a bone marrow aspirate, her physician was forced to use other methods to diagnose Gaucher's disease.

One year earlier she developed nonradiating epigastric pain. It was considered that this pain might be related to an intestinal virus disease, but when several of her associate nurses who also complained of abdominal pains improved, and she did not, she sought medical advice. Her physician was barely able to palpate the spleen and was unable to demonstrate abdominal tenderness. A liver–spleen radiographic scan was performed and splenic enlargement was demonstrated. She refused a bone marrow exam and leukocyte hydrolase assays were performed in February, 1979. White blood cell glucocerebrosidase and β-glucosidase activities were markedly reduced and if there has been a previously affected child in a family, the pregnancies at risk for these diseases can now be monitored. All nine of these lipid storage disorders are transmitted as recessive genetic abnormalities. In all but one the disease is carried on an autosomal chromosome. Fabry's disease is linked to the X chromosome. In all of these conditions one of four pregnancies statistically will be homozygous (or hemizygous in Fabry's disease), two fetuses will be carriers, and one will not be involved at all. The enzyme assay procedures have been used to detect affected fetuses and carriers in utero, using cultured fibroblasts obtained by amniocentesis as a source of enzyme.

There is no therapy for the sphingolipidoses; the role of medicine at the present time is prevention through genetic counseling based upon enzymologic assays of the type discussed above. A discussion of the diagnosis of Gaucher's disease is presented in Clin. Corr. 10.3.

10.5 THE PROSTAGLANDINS

Structure and Biosynthesis of Prostaglandins

The prostaglandins were discovered through their effects on smooth muscle, specifically their ability to promote the contraction of intestinal and uterine muscle and the lowering of blood pressure. Although complexity of their structures and the diversity of their sometimes conflicting functions sometimes create a sense of frustration, the potent pharmacological effects of the prostaglandins have afforded them an important place in human biology and medicine. With the exception of the red blood cell, the prostaglandins are produced and released by nearly all mammalian cells and tissues; they are not confined to specialized cells as insulin is to the pancreas. Furthermore, unlike most other hormones, the prostaglandins are not stored in cells but instead are synthesized and released immediately.

There are three major classes of primary prostaglandins, the A, E, and F series. The structures of the more common prostaglandins A, E, and F are shown in Figure 10.74. They are all related to prostanoic acid (Figure 10.75). Note that the prostaglandins contain a multiplicity of functional groups; for example, PGE_2 contains a carboxyl group, a β-hydroxyketone, a secondary alkylic alcohol and two carbon–carbon double bonds. The three classes (A, E, and F) are distinguished on the basis of the functional groups about the cyclopentane ring: the E type is a β-hydroxyketone, the F series are 1,3-diols, and those in the A series are α,β-unsaturated ketones. The

Glucocerebrosidase and β-Glucosidase Activities in Leukocytes from the Propositus

Source of Leukocytes	*β-Glucosidase Activity*[a] *[nmol/(hr · mg protein)]*	*Glucocerebrosidase Activity*[b] *[nmol/(hr · mg protein)]*
Controls (n = 25)[c]		
Range	6.50–13.8	4.45–7.67
Mean	8.42	5.59
Patients with confirmed type 1 Gaucher's disease		
A.A.	1.69	1.51
P.T.	1.65	0.94
T.P.	1.20	0.70
Propositus (H.V.)		
February, 1979	1.00[12][d]	0.80[14]
April, 1980	0.95[11]	0.94[17]

[a] Determined using the artificial, fluorogenic substrate 4-methylumbelliferyl-β-D-glucopyranoside in the presence of 1.2% (w/v) sodium taurocholate under conditions that measure relative glucocerebrosidase activity.

[b] Determined using authentic, radiolabeled glucocerebroside as the substrate.

[c] n, number of individual samples.

[d] The number in brackets represents the value for the propositus expressed as a percentage of the mean value for the controls.

were in the range of three other Gaucher patients. (see table). In addition, her serum acid phosphatase value was elevated more than four-fold. The elevated acid phosphatase activity was resistant to inhibition by L(+)-tartrate and had the electrophoretic mobility of a type 5 isoenzyme in acid (pH 4.3) polyacrylamide gels. She continued to suffer from mild abdominal discomfort but received no therapy. A liver–spleen scan was repeated one year later and a reduction in spleen size was noted. This reduction in size prompted her physicians to doubt the original diagnosis, since the spleen would not be expected to decrease in size if she had Gaucher's disease. White blood cell glucocerebrosidase and β-glucosidase determinations were repeated in April, 1980 and low values (less than 20% of control) of residual activity were found that were comparable to those of other patients with established Gaucher's disease. Her leukocyte glucocerebrosidase and β-glucosidase values had not changed significantly in the interval.

Because of the conflicting opinions the patient was urged to undergo a bone marrow exam and she consented. A bone marrow aspirate showed numerous Gaucher cells. Electron microscopy using negatively stained material that was scraped from a slide containing dried bone marrow showed twisted storage deposits typical of Gaucher's disease.

Over the next two months, her platelet count and hemoglobin concentration dropped significantly. Splenectomy was performed, her platelets and hemoglobin values returned to the normal range and she is doing well at this time.

subscript numerals 1, 2, or 3, refer to the number of double bonds in the side chains. All double bonds are of the trans configuration. The subscript "α" refers to the configuration of the C-9 hydroxyl group: an α-hydroxyl group projects "down" from the plane of the ring.

The most important dietary precursor of the prostaglandins is linoleic acid (18 : 2) which is an essential fatty acid. In adults linoleic acid is ingested daily in amounts of about 10 g. Only a very minor part of this total intake is converted by elongation and desaturation in the liver to arachidonic acid and to some extent also to dihomo-γ-linoleic acid. Since the total daily excretion of prostaglandins and their metabolites is only about 1 mg, it is clear that the formation of prostaglandins is a quantitatively unimportant pathway in the overall metabolism of fatty acids. At the same time, however, the metabolism of prostaglandins is completely dependent on a regular and

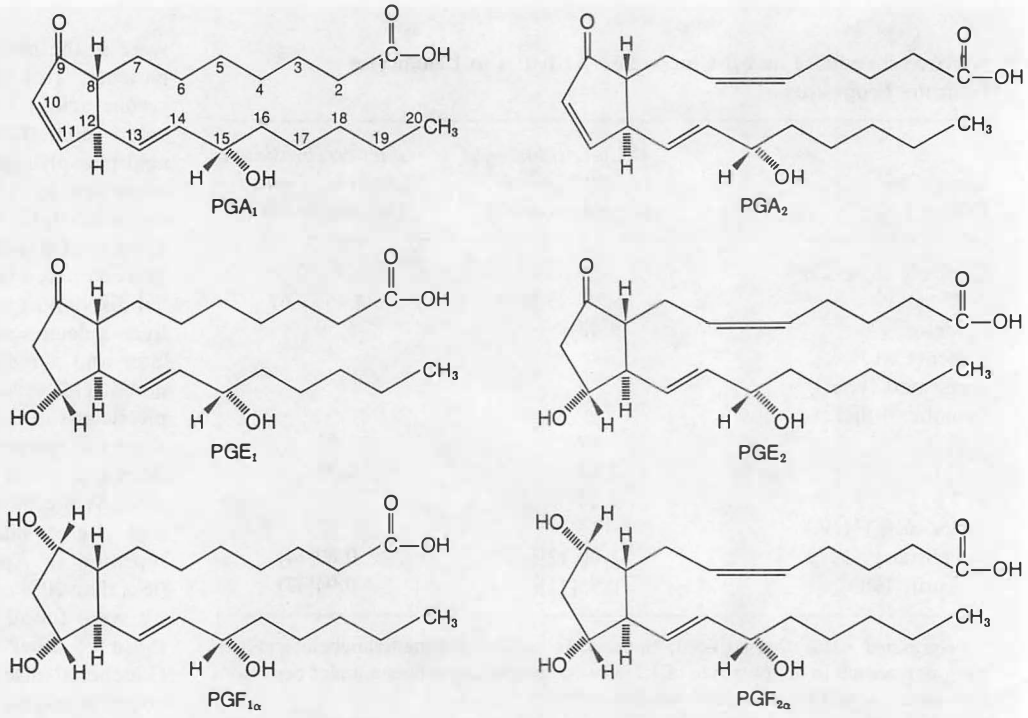


Figure 10.74
The structures of the major prostaglandins.

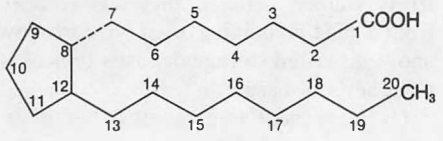


Figure 10.75
The structure of prostanoic acid.

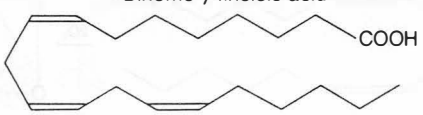
constant supply of linoleic acid. When the diet is deficient in linoleic acid, then there is decreased production of prostaglandins.

The immediate precursors to the prostaglandins are 20-carbon polyunsaturated fatty acids containing 3, 4, and 5 carbon–carbon double bonds. During their transformation into various prostaglandins they are cyclized and take up oxygen. Dihomo-γ-linoleic acid (C_{20}-Δ^8, 11,14) is the precursor to PGE_1 and $PGF_{1\alpha}$; arachidonic acid (C_{20}-Δ^5,8,11,14) is the precursor to PGE_2 and $PGF_{2\alpha}$; and eicosopentaenoic acid (C_{20}-20Δ^5,8,11,14,17) is the precursor to PGE_3 and $PGF_{3\alpha}$. (See Figure 10.76).

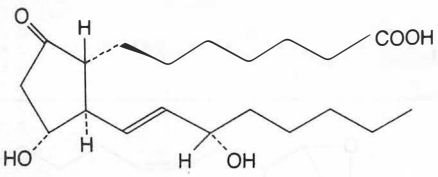
Compounds of the 2-series derived from arachidonic acid are the principal prostaglandins in man and are of the greatest significance biologically. Thus, one should focus attention primarily on the metabolism of arachidonic acid.

Figure 10.76
The synthesis of E and F prostaglandins from fatty acid precursors.

Dihomo-γ-linoleic acid

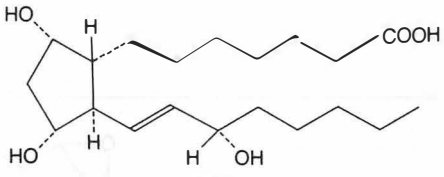


C_{20}-Δ8,11,14
(precursor)



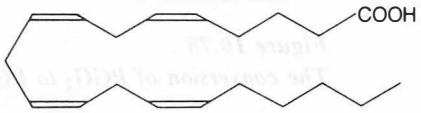
PGE_1

PGA_1, PGB_1, and their
19-hydroxy derivatives

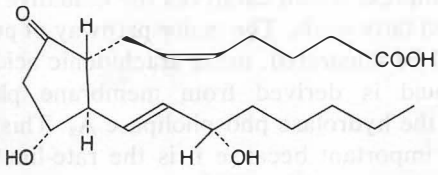


$PGF_{1\alpha}$

Arachidonic acid

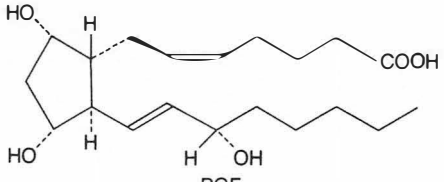


C_{20}-Δ5,8,11,14
(precursor)



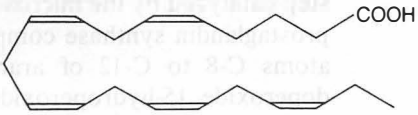
PGE_2

PGA_2, PGB_2, and their
19-hydroxy derivatives

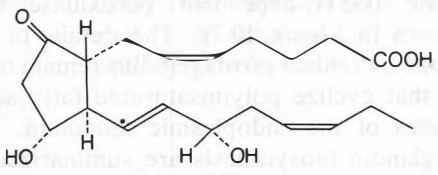


$PGF_{2\alpha}$

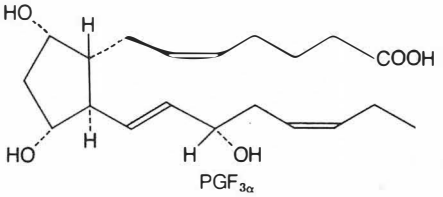
5,8,11,14,17-Eicosopentaenoic acid



C_{20}-Δ5,8,11,14,17
(precursor)



PGE_3



$PGF_{3\alpha}$

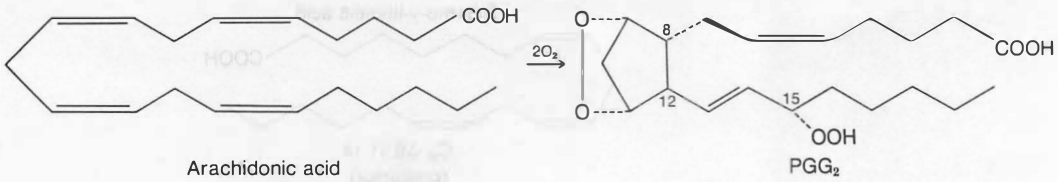


Figure 10.77
The cyclooxygenase reaction.

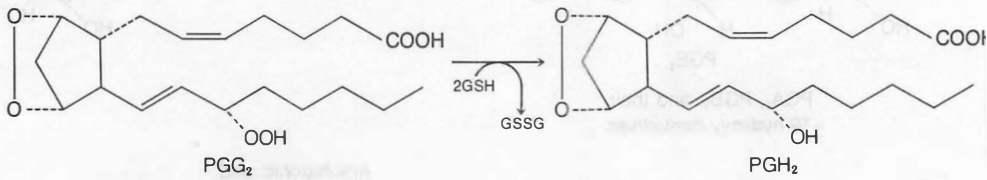


Figure 10.78
The conversion of PGG_2 to PGH_2.

The central enzyme system in prostaglandin biosynthesis is the prostaglandin synthase complex, which catalyzes the oxidative cyclization of polyunsaturated fatty acids. The major pathway of prostaglandin biosynthesis will be illustrated, using arachidonic acid as substrate. Arachidonic acid is derived from membrane phospholipids by the action of the hydrolase phospholipase A_2. This esterolytic cleavage step is important because it is the rate-limiting step in prostaglandin synthesis and because certain agents that stimulate prostaglandin production act by stimulating the activity of phospholipase A_2. Cholesterol esters containing arachidonic acid may also serve as a source of arachidonic acid substrate. The first step catalyzed by the microsomal cyclooxygenase component of the prostaglandin synthase complex involves the cyclization of carbon atoms C-8 to C-12 of arachidonic acid to form the cyclic endoperoxide 15-hydroperoxide, PGG_2, and the reaction requires 2 moles of molecular oxygen as shown in Figure 10.77. The 15-endoperoxide PGG_2 is then converted to prostaglandin H_2 (PGH_2) by a reduced glutathione (GSH)-dependent peroxidase that catalyzes the reaction shown in Figure 10.78. The details of the additional steps leading to the individual prostaglandins remain to be elucidated. The reactions that cyclize polyunsaturated fatty acids are found in the membranes of the endoplasmic reticulum. The major pathways of prostaglandin biosynthesis are summarized in Figure 10.79. The formation of the primary prostaglandins of the D,

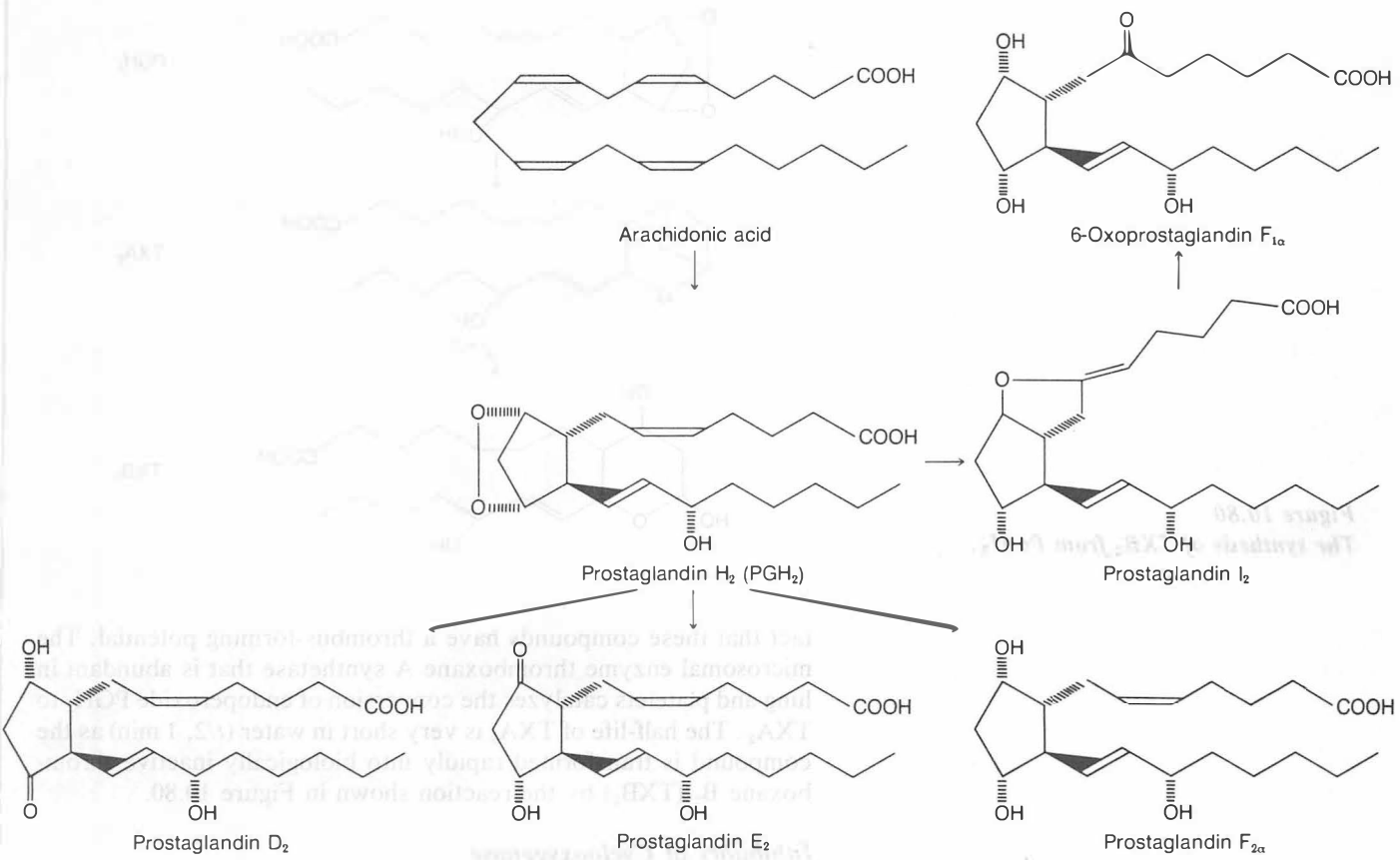


Figure 10.79
Major routes of prostanoid biosynthesis.

E, and F series into thromboxanes or into prostacyclin (PGI_2) is mediated by different specific enzymes, whose presence varies depending upon the cell type and tissue. This results in a degree of tissue specificity as to the type and quantity of prostaglandin produced. Thus, in the kidney and spleen PGE_2 and $PGF_{2\alpha}$ are the major prostaglandins produced. In contrast, blood vessels produce mostly PGI_2. In the heart PGE_2, $PGF_{2\alpha}$, and PGI_2 are formed in about equal amounts. Thromboxane A_2 (TXA_2) is the main prostaglandin endoperoxide formed in platelets.

The prostaglandins have a very short half-life. Soon after release they are rapidly taken up by cells and inactivated. The lungs appear to play an important role in inactivating prostaglandins.

Thromboxanes, mentioned above, are the highly active metabolites of the PGG_2- and PGH_2-type prostaglandin endoperoxides that have the cyclopentane ring replaced by a six-membered oxygen-containing (oxane) ring. The term thromboxane is derived from the

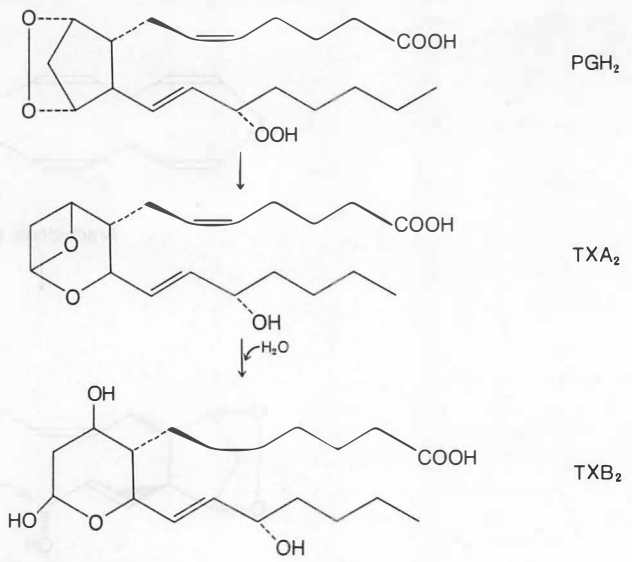


Figure 10.80
The synthesis of TXB_2 from PGH_2.

fact that these compounds have a thrombus-forming potential. The microsomal enzyme thromboxane A synthetase that is abundant in lung and platelets catalyzes the conversion of endoperoxide PGH_2 to TXA_2. The half-life of TXA_2 is very short in water ($t/2$, 1 min) as the compound is transformed rapidly into biologically inactive thromboxane B_2 (TXB_2) by the reaction shown in Figure 10.80.

Inhibitors of Cyclooxygenase

Clinically there are two types of drugs that affect prostaglandin metabolism and are therapeutically useful. First, there are the nonsteroidal, antiinflammatory agents such as aspirin (acetylsalicylic acid), indomethacin, and phenylbutazone, which block prostaglandin production by irreversibly inhibiting the enzyme cyclooxygenase. In the case of aspirin, inhibition occurs presumably by acetylation of the enzyme. These drugs are not without their undesirable side effects: aplastic anemia can result from phenylbutazone therapy. The second group, the steroidal antiinflammatory drugs like hydrocortisone, prednisone, and betamethasone, appear to act by blocking prostaglandin release by inhibiting phospholipase A_2 activity so as to interfere with mobilization of arachidonic acid, the substrate for cyclooxygenase. (See Figure 10.81.)

The factors that govern the biosynthesis of prostaglandins are poorly understood, but, in general, prostaglandin release seems to be triggered following hormonal or neural excitation or after muscular activity. For example, histamine stimulates an increase in the

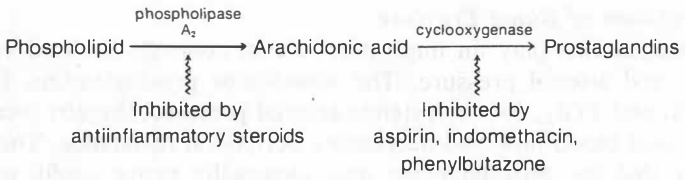


Figure 10.81
Site of action of inhibitors of prostaglandin synthesis.

prostaglandin concentration in gastric perfusates. Also, prostaglandins are released during labor and after cellular injury (e.g., platelets exposed to thrombin, lungs irritated by dust).

Physiological Effects of the Prostaglandins

Inflammation

Prostaglandins appear to be one of the natural mediators of inflammation. Inflammatory reactions most often involve the joints (rheumatoid arthritis), skin (psoriasis), and eyes, and inflammation of these sites is frequently treated with corticosteroids that inhibit prostaglandin synthesis. Administration of the prostaglandins PGE_2 and PGE_1 induce the signs of inflammation that include redness and heat (due to arteriolar vasodilation), and swelling and edema resulting from increased capillary permeability.

Pain and Fever

PGE_2 in amounts that alone do not cause pain, prior to administration of the autocoids, histamine and bradykinin, enhance both the intensity and the duration of pain caused by these two agents. It is thought that pyrogen activates the prostaglandin biosynthetic pathway, resulting in the release of PGE_2 in the region of the hypothalamus where body temperature is regulated. Aspirin, which is an antipyretic drug, acts by inhibiting cyclooxygenase.

Reproduction

The prostaglandins have been used extensively as drugs in the reproductive area. PGF_2 and PGE_2 have been used to induce parturition and for the termination of an unwanted pregnancy. There is also evidence that the PGE series of prostaglandins may play some role in infertility in males.

Gastric Secretion and Peptic Ulcer

Synthetic prostaglandins have proven to be very effective in inhibiting gastric acid secretion in patients with peptic ulcers. The inhibitory effect of PGE compounds appears to be due to inhibition of cAMP formation in gastric mucosal cells. Prostaglandins also accelerate the healing of gastric ulcers.

Regulation of Blood Pressure

Prostaglandins play an important role in controlling blood vessel tone and arterial pressure. The vasodilator prostaglandins, PGE, PGA, and PGI_2, lower systemic arterial pressure, thereby increasing local blood flow and decreasing peripheral resistance. There is hope that the prostaglandins may eventually prove useful in the treatment of hypertension.

Ductus Arteriosus and Congenital Heart Disease

PGE_2 functions in the fetus to maintain the patency of the ductus arteriosus prior to birth. There are two clinical applications of prostaglandin biochemistry in this area. First, if the ductus remains open after birth, closure can be hastened by administration of the cyclooxygenase inhibitor indomethacin. In other situations it may be desirable to keep the ductus open. For example, in the case of infants born with congenital abnormalities where the defect can be corrected surgically, infusion of prostaglandins will maintain blood flow through the ductus over this interim period.

Platelet Aggregation and Thrombosis

Certain prostaglandins, especially PGI_2, inhibit platelet aggregation, whereas PGE_2 and TXA_2 promote this clotting process. TXA_2 is produced by platelets and accounts for the spontaneous aggregation that occurs when platelets contact some foreign surface, collagen or thrombin. Endothelial cells lining blood vessels release PGI_2 and may account for the lack of adherence of platelets to the healthy blood vessel wall.

BIBLIOGRAPHY

Phospholipid Metabolism

Holub, B. J., and Kuksis, A. Metabolism of molecular species of diacylglycerophospholipids. *Adv. Lipid Res.* 16:1, 1978.

Sphingolipids and the Sphingolipidoses

Glew, R. H., and Peters, S. P. *Practical enzymology of the sphingolipidoses*. New York: Alan R. Liss, 1977.

Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. *The metabolic basis of inherited disease*, 4th ed. New York: McGraw-Hill, 1978, p. 718.

Lung Surfactant

Shelly, S. A., Kovacevic, M., Paciga, J. E., and Balis, J. U. Sequential changes of surfactant phosphatidylcholine in hyaline-membrane disease of the newborn. *N. Engl. J. Med.* 300:112, 1979.

Torday, J., Carson, L., and Lawson, E. E. Saturated phosphatidylcholine in amniotic fluid and prediction of the respiratory distress syndrome. *N. Engl. J. Med.* 300:1013, 1979.

Prostaglandins

Curtis-Prior, B. P. *Prostaglandins, an introduction to their biochemistry, physiology and pharmacology*. New York: North-Holland, 1976.

Samuelsson, B. Prostaglandins and thromboxanes, in R. O. Greep (ed.), *Recent progress in hormone research*, vol. 34. New York: Academic, 1978, p. 239.

Cholesterol Synthesis

Brown, M. S., and Goldstein, J. L. Receptor-mediated control of cholesterol metabolism. *Science* 191:150, 1976.

Small, D. M. Cellular mechanisms for lipid deposition in arteriosclerosis. *N. Engl. J. Med.* 297:873 and 924, 1977.

Bile Acids

Hanson, R. F., and Pries, J. M. Synthesis and enterohepatic circulation of bile salts. *Gastroenterology* 73:611, 1977.

11

Amino Acid Metabolism I

THOMAS I. DIAMONDSTONE

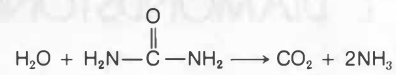
GENERAL PATHWAYS

11.1 OVERVIEW

Unlike fats, proteins taken in by the animal body in excess of immediate need cannot be stored as such; instead, they are metabolized in some fashion. In addition, there is an ongoing breakdown of endogenous protein, which must be continually replenished. Quantitatively, it has been estimated that in a 70-kg man on a normal diet, something of the order of 400 g of protein turns over daily, of which up to one-fourth is oxidized. During the course of this oxidation, nitrogen from the amino acids making up the proteins is removed as ammonia.

Although ammonia in small amounts is a normal metabolite, in larger quantities it becomes dangerous. Most of it is detoxified in the liver by conversion into the nontoxic substance *urea* (Figure 11.1) and excreted in the urine; the detoxification of ammonia is one of the major tasks carried out by the liver in the metabolism of protein amino acids. Some ammonia is incorporated into glutamine, a reaction catalyzed by *glutamine synthetase* (Chapter 12). This reaction provides a pool of nitrogen for synthetic reactions, which is both nontoxic and at the reduction level of ammonia. In the kidney glutamine provides a source of readily excretable ammonia (as NH_4^+) as part of a mechanism for maintaining acid–base balance (Chapter 23).

Blood levels of ammonia average ~100 μg/100 ml, except in the portal vein, where the concentration is ~350 μg/100 ml. The source of this ammonia is urea, which, being freely diffusible, finds its way into the large intestine, where it is acted upon by bacterial urease:



The ammonia produced in this reaction is incorporated into glutamine.

For purposes of discussion, amino acid metabolism can be separated into two main parts: (1) disposal of amino nitrogen, discussed in this chapter, and (2) the fate of the carbon skeletons. This latter aspect will be discussed in detail in the next chapter; however, in order to put the concepts and information in context, it will be noted here that in general the carbon skeletons can be directly oxidized to carbon dioxide and water or be converted back to glucose and then to carbon dioxide and water. In addition, some amino acids can serve as precursors for other biologically important compounds, such as purines, pyrimidines and neurotransmitters.

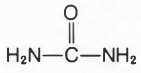


Figure 11.1
Structure of urea.

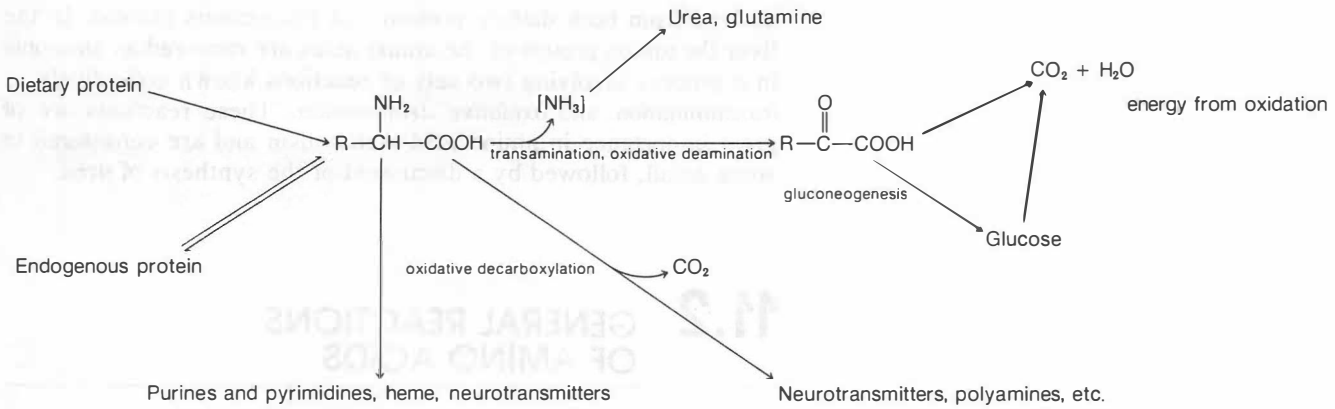


Figure 11.2
Overall metabolism of protein amino acids.

A summary scheme of the overall metabolism of protein amino acids is shown in Figure 11.2. As shown in the figure, amino acids which serve as precursors to polyamines and some neurotransmitters first undergo decarboxylation. The individual decarboxylation reactions will be discussed in Chapter 12 in the contexts in which they occur.

The digestion of dietary proteins begins in the stomach and takes place mainly in the small intestine. In these organs the polypeptide chains are hydrolyzed to the individual amino acids by the action of a variety of proteolytic enzymes. (See Chapter 24 for a detailed discussion of the digestion of proteins.) The amino acids thus generated are carried to the liver via the superior mesenteric and portal veins. There is evidence that absorption of amino acids from the small intestine into the circulation is greatly dependent on pyridoxal phosphate (Figure 11.3), the active form of vitamin B_6. Indeed, this compound appears to play a significant role in the uptake of amino acids by cells in general.

The pool of amino acids supplied to the liver consists of those

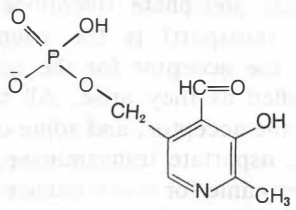


Figure 11.3
Structure of pyridoxal phosphate.

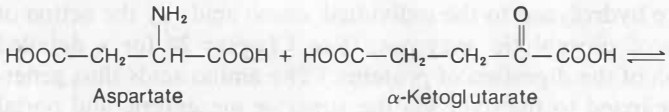
derived from both dietary protein and endogenous protein. In the liver the amino groups of the amino acids are removed as ammonia in a process involving two sets of reactions known collectively as *transamination* and *oxidative deamination*. These reactions are of great importance in amino acid metabolism and are considered in some detail, followed by a discussion of the synthesis of urea.

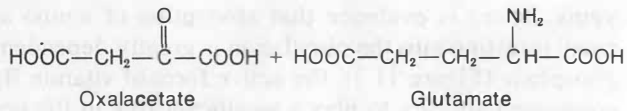
11.2 GENERAL REACTIONS OF AMINO ACIDS

Transamination

This extremely widespread set of reactions constitutes the major mechanism by which protein amino acids are divested of their amino groups; virtually every protein amino acid undergoes transamination at some point in its metabolic breakdown. The variety of individual transamination reactions will become apparent in Chapter 12, during the discussion of the metabolism of the individual amino acids.

Transamination involves the transfer of an amino group from one carbon skeleton to another, as illustrated in the reaction catalyzed by *aspartate transaminase:*





In the reaction as written, aspartate is the donor amino acid and α-ketoglutarate is the acceptor keto acid. For all the known transaminases, pyridoxal phosphate (mentioned above in connection with amino acid transport) is the coenzyme. For nearly all, α-ketoglutarate is the acceptor for the amino group. The exceptions will be handled as they arise. All transaminases for which α-ketoglutarate is the acceptor, and some others, are named by the donor amino acid: aspartate transaminase, tyrosine transaminase, and so on. Another name for transaminase is *aminotransferase,* and this name is also widely used.

Cytoplasmic aspartate transaminase from pig heart has been the

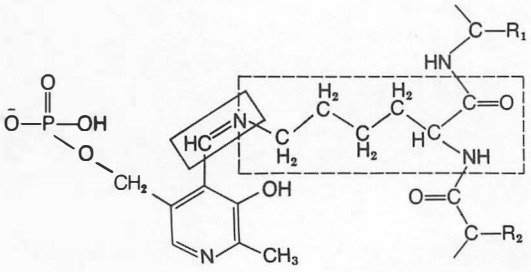


Figure 11.4
The aldimine linkage between the aldehyde group of pyridoxal phosphate and the ϵ-amino group of lysine (solid rectangle).
The dashed rectangle indicates the lysine residue in the polypeptide chain.

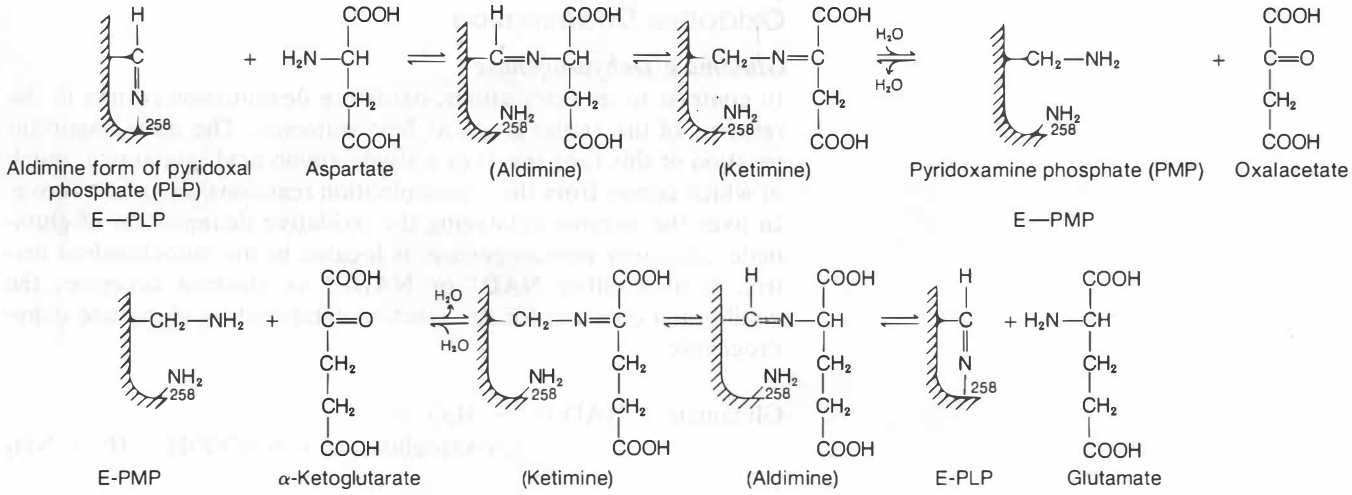


Figure 11.5
The mechanism of enzymatic transamination represented in schematic form.
The interconversions between aldimine and ketimine appear to be rate-limiting in both directions.

most extensively studied transaminase, and from this work has come most of our knowledge of the mechanism of enzymatic transamination. The following is a summary of this information.

When the enzyme is isolated from tissue and purified, the coenzyme is bound to it. The coenzyme exists in the aldehyde form, and the aldehyde functional group is linked to the ε-amino group of lysine-258 as an aldimine or Schiff base (Figure 11.4). Aspartate reacts with this internal Schiff base, generating oxalacetate and the amine form of the bound coenzyme. The pyridoxamine form of the coenzyme then reacts with α-ketoglutarate, producing glutamate and regenerating the lysine-pyridoxal phosphate Schiff base. This set of reactions is shown in Figure 11.5. (The phosphate group, not shown in the figure, is an absolute requirement for activity.) These reactions are freely reversible; equilibrium constants for transaminations are in the range of 1 to 10.

Kinetically, each substrate (amino donor, keto acceptor) reacts in turn with an enzyme form to generate a product, the amino donor giving rise to the product keto acid and the keto acceptor becoming the product amino acid; the enzyme oscillates between aldehyde and amine forms. This reaction mechanism, in which each substrate reacts in turn with a different enzyme form to yield a single product, is characteristic of the so-called ping-pong mechanism (see Chapter 4).

Oxidative Deamination

Glutamate Dehydrogenase

In contrast to transamination, oxidative deamination results in the removal of the amino group as free ammonia. The most important reaction of this type involves a single amino acid, glutamate, much of which comes from the transamination reactions described above. In liver the enzyme catalyzing the oxidative deamination of glutamate, *glutamate dehydrogenase,* is located in the mitochondrial matrix. It uses either NAD^+ or $NADP^+$ as electron acceptor; the equilibrium constant for the reaction catalyzed by glutamate dehydrogenase,

$$\text{Glutamate} + \text{NAD(P)}^+ + \text{H}_2\text{O} \rightleftharpoons \alpha\text{-ketoglutarate} + \text{NAD(P)H} + \text{H}^+ + \text{NH}_3$$

lies far in the direction of glutamate synthesis. Thus, at first glance, this reaction would seem to be useless for the purpose of generating ammonia to be used in urea synthesis. This apparent paradox will be resolved shortly.

As in any reaction involving the reduction of the pyridine nucleotides, the electrons transferred to NAD (or NADP) are ultimately delivered to oxygen via the electron transport chain (Chapter 6):

$$\text{NAD(P)H} + \text{H}^+ + \tfrac{1}{2}\text{O}_2 \longrightarrow \text{NAD(P)}^+ + \text{H}_2\text{O}$$

so that the overall reaction can be written

$$\text{Glutamate} + \tfrac{1}{2}\text{O}_2 \longrightarrow \alpha\text{-ketoglutarate} + \text{NH}_3$$

In common with other dehydrogenases, the kinetic mechanism of the glutamate dehydrogenase reaction is ordered in both directions, with the pyridine nucleotide adding first.

Glutamate dehydrogenase exists in a number of forms. Its basic form is that of an oligomer composed of six identical polypeptide chains, which are capable of binding GTP and ADP as well as substrates. It is interesting to note that the enzyme can exist in a form in which it exhibits dehydrogenase activity toward alanine and other monocarboxylic acids, although with a much lower specific activity. The combination of GTP and NADH promote this activity, while at the same time suppressing glutamate dehydrogenase activity; ADP has the opposite effect. Some of the possible implications of this regulatory control are discussed below.

Amino Acid Oxidases

A relatively minor pathway for removal of amino groups from amino acids involves the L- and D-*amino acid oxidases,* found in the peroxisomes of liver and kidney. The immediate electron acceptor in these reactions is a tightly bound flavin (FMN for the L-amino acid oxidase, FAD for the D-amino acid oxidase). The reaction is

$$\text{Amino acid} + H_2O + E{-}F \longrightarrow \text{keto acid} + NH_3 + E{-}FH_2$$

where E—F and E—FH_2 represent, respectively, the oxidized and reduced forms of the enzyme-bound flavins. The reduced form, E—FH_2, can react directly with molecular oxygen within the peroxisomes:

$$E{-}FH_2 + O_2 \longrightarrow E{-}F + H_2O_2$$

regenerating the oxidized form of the enzyme and producing hydrogen peroxide, which in turn is decomposed to water and oxygen by the action of catalase:

$$H_2O_2 \longrightarrow H_2O + \tfrac{1}{2}O_2$$

Thus, the overall reaction is

$$\text{Amino acid} + \tfrac{1}{2}O_2 \longrightarrow \text{keto acid} + NH_3$$

exactly analogous to the reaction catalyzed by glutamate dehydrogenase.

The specificities of the two oxidases are rather broad. L-Amino acid oxidase activity is low, in contrast to the relatively high activity of D-amino acid oxidase, especially in kidney and brain; the physiological role of both enzymes is unclear. It has been suggested that D-amino acids arise from the breakdown of intestinal bacteria. This enzyme also catalyzes the oxidation of glycine and sarcosine (*N*-methylglycine), generating glyoxylate and either ammonia or methylamine:

$$RHN{-}CH_2{-}COOH \longrightarrow OHC{-}COOH + RNH_2$$

R = H (glycine)
R = CH_3 (sarcosine)

However, this is not a major route of glycine metabolism. The rather high level of D-amino acid oxidase in brain has led to the suggestion

that it may be important in relation to the function of glycine as a neurotransmitter.

Glutamate Dehydrogenase and the Flow of Nitrogen

As stated above, the reactions catalyzed by D- and L-amino acid oxidases contribute only a small fraction of the total ammonia pool. The great bulk of ammonia comes via the combination of transamination and the reaction catalyzed by glutamate dehydrogenase; these two reactions serve to "funnel" ammonia into urea formation (Figure 11.6). Thus the flow of nitrogen is from the amino acids to urea, and this accounts for the fact that the glutamate dehydrogenase reaction can be run in a thermodynamically unfavorable direction.

By virtue of its ability to alter its substrate specificity in response to changes in the energy charge and reduction state of the cell (see

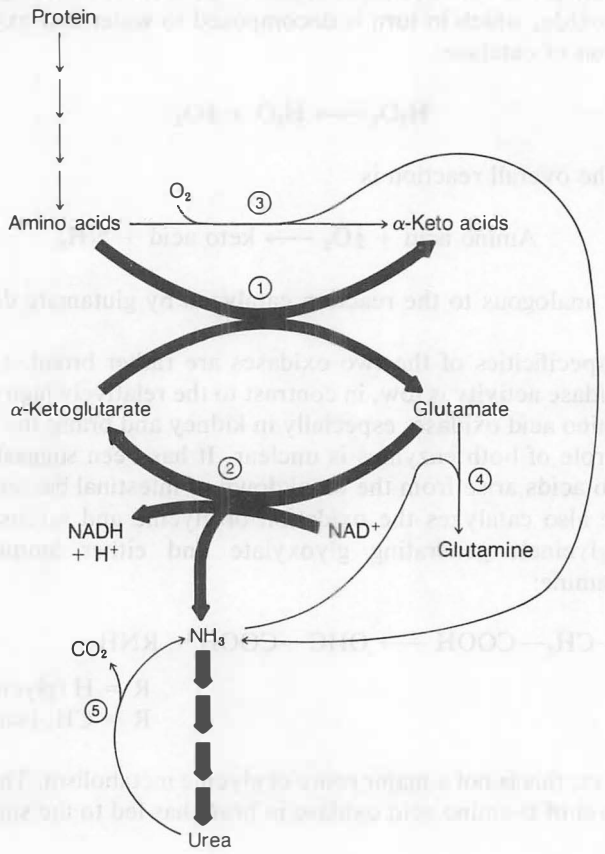


Figure 11.6
Disposal of amino nitrogen, emphasizing transamination of amino and deamination of glutamate.
The enzymes are (1) specific amino acid transaminases, (2) glutamate dehydrogenase, (3) amino acid oxidases, (4) glutamine synthetase, and (5) bacterial urease.

above), glutamate dehydrogenase can, to a certain extent, affect the overall rate at which the TCA cycle operates by influencing the level of α-ketoglutarate. With high ADP levels, reflecting an energy drain, more α-ketoglutarate is produced, whereas with high GTP and NADH levels (the latter perhaps arising from fatty acid oxidation), much less α-ketoglutarate is generated. Since the specific activity of the alanine dehydrogenase activity is some 1 to 2% of that of the glutamate dehydrogenase activity, the main effect of these changes is on the level of α-ketoglutarate.

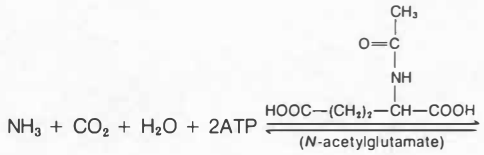
Concomitant with the changes in α-ketoglutarate production is a change in the amount of ammonia released (ultimately) from protein amino acids. Presumably, one of the responses to a lowered energy charge (high ADP) would be an increased degradation of protein amino acids to be used in gluconeogenesis (Chapter 12); the opposite would be true with a high energy charge (high GTP), which would lower considerably the rate at which protein amino acids are deaminated.

It should be noted that, whereas the formation of urea is essentially irreversible (except for its decomposition by bacterial urease), the formation of glutamine is not. Its breakdown is not shown in Figure 11.6, because it is metabolized in a variety of ways. Glutamine is the transport form of ammonia, and in many parts of the body its formation is the first reaction in which ammonia participates. The formation of glutamine is the only ammonia detoxification reaction available to brain tissue.

11.3 THE UREA CYCLE

Having discussed the formation of ammonia from amino acids, we are now in a position to consider its disposal as urea, the primary nitrogenous constituent of urine. Urea accounts for about 80% by weight of these constituents under conditions of normal protein intake. The absolute amount of urea nitrogen ranges ~6–17 g/day, which corresponds to ~35–100 g of protein. Almost all of the urea produced in this fashion is excreted in the urine.

Urea is of considerable interest in the history of both biochemistry and organic chemistry. It was the first substance of biological origin to be synthesized in the laboratory; this was accomplished in 1828 by Friedrich Wöhler, who obtained urea by boiling an aqueous solution of ammonium cyanate (NH_4OCN). The biological synthesis of urea is carried out by a cyclic set of reactions known variously as the



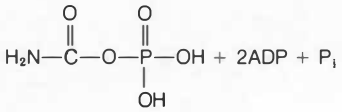


Figure 11.7
Formation of carbamoyl phosphate.
Note the requirement for 2 moles of ATP, and also the requirement for N*-acetylglutamate.*

urea cycle, the *ornithine cycle*, or the *Krebs–Henseleit cycle* (this last after its discoverers). In this chapter it will be referred to as the *urea cycle*. This cycle of reactions, quantitatively most important in the liver, is also present in kidney, skin, brain, and blood cells. In tissues other than liver it appears to serve primarily as a mechanism for synthesizing arginine.

Reactions of the Cycle

The first reaction involved in the synthesis of urea is the formation of carbamoyl phosphate. This reaction is not part of the urea cycle proper; rather, it is an "entry" reaction, just as the formation of acetyl CoA is an "entry" reaction into the TCA cycle, and its function is to supply the carbamoyl ($H_2N—\overset{O}{\overset{\|}{C}}—$) group.

The formation of carbamoyl phosphate (Figure 11.7) requires CO_2, 2 moles of ATP, and the ammonia released in the glutamate dehydrogenase reaction; the products are carbamoyl phosphate, inorganic phosphate (P_i), and ADP. It takes place in the mitochondrial matrix and is catalyzed by *carbamoyl phosphate synthetase (ammonia)*, also known as *carbamoyl phosphate synthetase I*. This enzyme is different from *carbamoyl phosphate synthetase (glutamine)*, a superficially similar cytosolic enzyme involved in the biosynthesis of pyrimidines (Chapter 13). The mitochondrial enzyme exhibits an absolute requirement for *N*-acetylglutamate, a positive allosteric effector.

In the next step (Figure 11.8) the carbamoyl group is transferred to the δ-amino group of ornithine; this reaction is the first in the urea cycle proper. As in the formation of carbamoyl phosphate, this reaction takes place in the mitochondrial matrix. The product is citrulline, and the reaction is catalyzed by *ornithine transcarbamylase;*

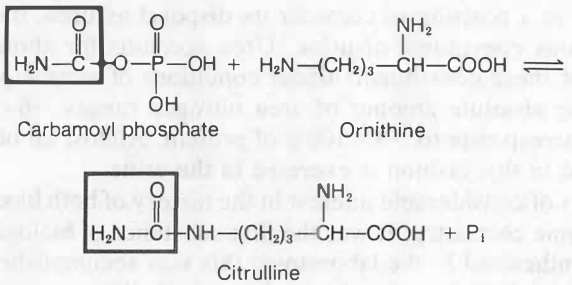


Figure 11.8
Formation of citrulline.

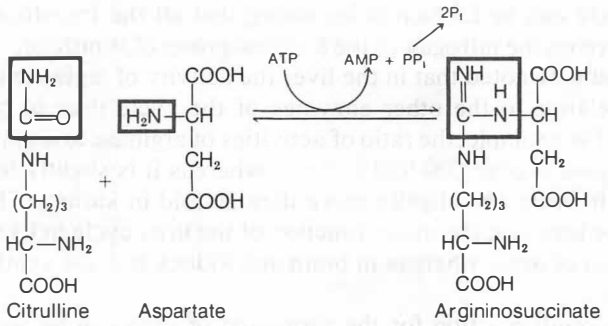


Figure 11.9
Formation of argininosuccinate.
Note the production of AMP and inorganic pyrophosphate.

concomitantly, a mole of inorganic phosphate is released. The citrulline thus produced leaves the mitochondrion by passive diffusion and enters the cytosol, where the remaining reactions of the urea cycle take place.

In the third reaction, catalyzed by *argininosuccinate synthetase*, the second nitrogen of urea is acquired from aspartate; it combines with citrulline in the presence of ATP so that the amino nitrogen of aspartate becomes covalently linked to the carbamoyl carbon of citrulline, as shown in Figure 11.9. The initial reaction is probably the formation of a Schiff base ($>C{=}N-$) between citrulline and aspartate. The products are argininosuccinate, AMP, and inorganic pyrophosphate (PP_i) (a strong inhibitor of the reaction). The cleavage of pyrophosphate to orthophosphate is catalyzed by inorganic pyrophosphatase, large quantities of which are present in liver and kidney. This cleavage supplies additional energy for the synthetic reaction, and it also removes the inhibitory effect of pyrophosphate. The main source of aspartate is cytoplasmic oxaloacetate, which is converted to aspartate via transamination; however, a small portion comes directly from protein. Argininosuccinate is cleaved (Figure 11.10) in the presence of *argininosuccinate lyase* to produce arginine and fumarate; the latter is converted to malate by an extramitochondrial fumarase.

Up to this point the pathway to arginine is common to all organisms that synthesize arginine. Ureotelic (urea-excreting) organisms, including humans, possess *arginase*, which catalyzes the splitting of arginine into urea and ornithine, as shown in Figure 11.11. Urea, of course, is the desired end product, and the release of ornithine completes the cycle. The ornithine reenters the mitochondrion through the agency of a specific transport mechanism. The entire cycle of reactions is shown in Figure 11.12.

Understanding of the apparent complexity of the reactions of the

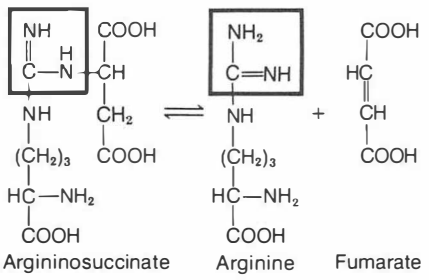


Figure 11.10
Formation of arginine.

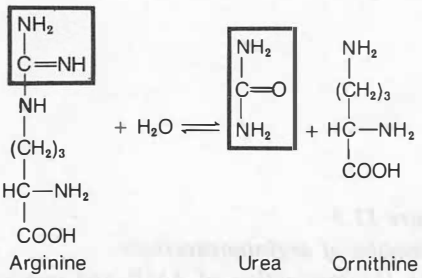


Figure 11.11
Formation of urea.

urea cycle can be facilitated by noting that all the transformations take place on the nitrogen of the δ-amino group of ornithine.

It should be noted that in the liver the activity of arginase is much larger relative to the other enzymes of the cycle than in brain or kidney. For example, the ratio of activities of arginase to argininosuccinate lyase is over 200-fold in liver, whereas it is slightly less than 30-fold in brain and slightly more than 20-fold in kidney. This supports the idea that the major function of the urea cycle in liver is the formation of urea, whereas in brain and kidney it is the synthesis of arginine.

The overall reaction for the formation of urea can be written in different ways, depending on what one wishes to emphasize. For example, during the course of the operation of the cycle one molecule of urea nitrogen comes from ammonia and the other from aspartate; thus the reaction could be written

$$\text{Aspartate} + NH_3 + CO_2 + 3\text{ATP} \longrightarrow 3H_2O + \text{urea} + \text{fumarate} + 2\text{ADP} + \text{AMP} + 4P_i$$

This formulation also points up the interconnection with the TCA cycle (see below).

Alternatively, the overall reaction could be written

$$2\ \text{Glutamate} + CO_2 + 3\text{ATP} + \text{NAD(P)} + 3H_2O \longrightarrow \text{urea} + 2\text{ADP} + \text{AMP} + 4P_i + \text{NAD(P)H} + H^+ + 2\ \alpha\text{-ketoglutarate}$$

Writing the reaction in this way shows that both nitrogen atoms ultimately derive from glutamate; the one appearing as ammonia coming from the oxidative deamination of glutamate, the other coming from glutamate via transamination with oxalacetate to produce aspartate.

Relationship to the TCA Cycle

In a purely formal sense, the urea cycle could operate independently of the TCA cycle; all the reactions involving the interconversion of fumarate back to aspartate can be carried out in the cytosol. These reactions are, in order, hydration of fumarate to malate, oxidation of malate to oxalacetate, and transamination of oxalacetate to aspartate. In practice, however, interconnections do exist. The malate derived from fumarate probably does not stay in the cytosol; rather, it returns to the mitochondria in a specific exchange reaction.

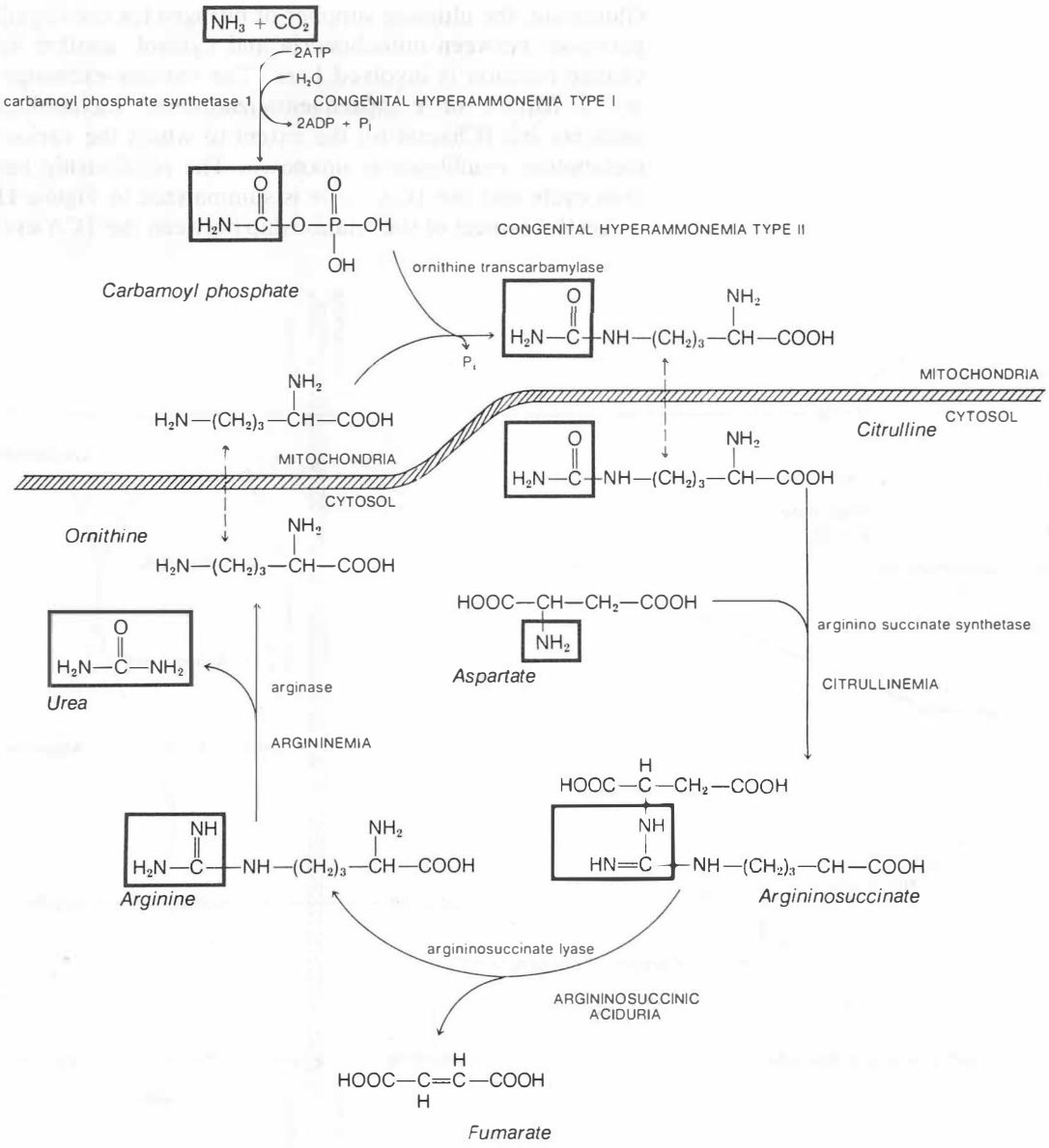


Figure 11.12
The urea cycle.
Names of compounds are in italic and names of clinical syndromes corresponding to deficiencies of particular enzymes in capitals.

Glutamate, the ultimate supplier of nitrogen for urea synthesis, also partitions between mitochondria and cytosol; another specific exchange reaction is involved here. The various exchange reactions are a feature of compartmentalization of metabolites by the mitochondria (Chapter 6); the extent to which the various pools of metabolites equilibrate is unknown. The relationship between the urea cycle and the TCA cycle is summarized in Figure 11.13.

Another aspect of the relationship between the TCA cycle and the

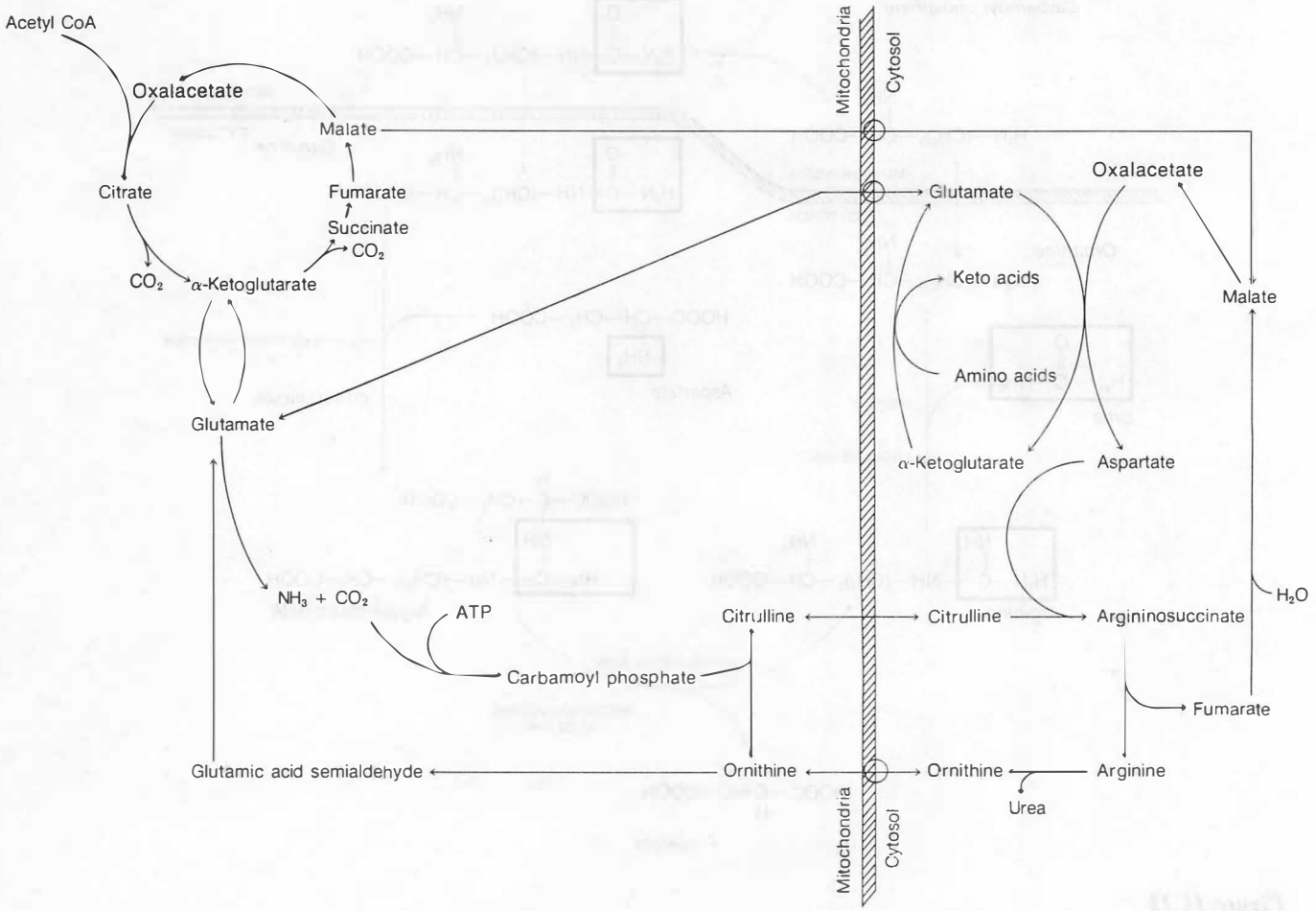


Figure 11.13
Interconnections between the urea cycle and the TCA cycle.
The formation of glutamic acid semialdehyde from ornithine is discussed in Chapter 12.

urea cycle has to do with the regulation of glutamate dehydrogenase (page 550). When the energy charge is high, or there is sufficient carbohydrate, relatively little degradation of amino acids and release of ammonia occur. When there is a large demand for protein amino acids, it is likely that the relative importance of the individual reactions of the TCA cycle shift. The rate of formation of citrate and isocitrate is probably diminished, whereas the rates of the other reactions are maintained or possibly increased, with a good share of oxalacetate being converted to phospho*enol*pyruvate. (See Chapter 12 for a discussion of gluconeogenesis from protein amino acids.) This regulation of glutamate dehydrogenase may be partly responsible for the protein-sparing effect of carbohydrate (i.e., the ability of carbohydrate to reduce the amount of dietary protein required to maintain nitrogen balance).

Regulation of the Urea Cycle

As noted above, urea accounts for about 80% of the total nitrogenous constituents of urine under conditions of normal protein intake. However, urea synthesis is markedly reduced when the diet contains little or no protein (i.e., essentially all carbohydrate); under these conditions, urea makes up only about 60% of the nitrogenous substances in urine and the levels of the urea cycle enzymes decrease. In contrast, increasing the protein intake above the normal level results in a concomitant increase of as much as threefold over that observed under normal protein intake for all the enzymes except for arginase, which increases by about 50%. In starvation, where the body is forced to use its own tissues as fuel, there is an even higher increase in enzyme levels, about four- to five-fold (about two-fold for arginase), owing to vastly enhanced gluconeogenesis from amino acids resulting from breakdown of tissue proteins (Chapter 12). Overall, variation of dietary protein intake from low to high to starvation produces relative changes in activity over a 10- to 20-fold range.

In addition to the coordinated response to dietary fluctuations by the urea cycle enzymes as a group, carbamoyl phosphate synthetase is controlled by its allosteric effector, *N*-acetylglutamate (Figure 11.7). This compound is synthesized from glutamate and acetyl CoA; the reaction is catalyzed by *acetylglutamate synthetase*. The activity of this enzyme is increased in the presence of arginine; this effect is maximal at an arginine concentration of ~0.1 mM in vitro. The effect of arginine seems to be to promote the dimerization of the enzyme into an aggregate of mol w 200,000. The mechanism by which this aggregation influences the activity of the enzyme is not yet known.

CLIN. CORR. 11.1 THE HYPERAMMONEMIC SYNDROME AND HEPATIC COMA: DISEASES ARISING FROM INBORN ERRORS OF UREA CYCLE ENZYMES

Hyperammonemia, or *hyperammonemic syndrome*, simply means elevated blood ammonia concentrations. This metabolic abnormality predisposes to coma, broadly speaking, although comatose states can be induced by other conditions besides hyperammonemia. (In fulminant hepatic failure, for example, coma ensues in the presence of essentially normal blood concentrations of ammonia.) Since elevated blood levels of ammonia point to disturbances of liver function, the resulting coma is often referred to as *hepatic coma*.

ACQUIRED HYPERAMMONEMIA: PORTAL–SYSTEMIC ENCEPHALOPATHY

Almost invariably, acquired hyperammonemia is a result of the development of collateral circulation around the liver in response to cirrhosis. The development of this alternate path of blood flow is known as *portal–systemic shunting* because the blood from the portal vein is diverted around the liver directly into the inferior vena cava. For this reason, the name *portal–systemic encephalopathy* (PSE) has been proposed for the constellation of neurologic, neuromuscular, and psychological abnormalities consequent upon acquired hyperammonemia.

When shunting develops, circulatory access to the liver is severely impaired, and as a result detoxification of ammonia is drastically reduced. Because of this,

ammonia arising from the extrahepatic metabolism of protein amino acids (plus some other sources), which now becomes quantitatively more important in the nitrogen economy, cannot be converted to urea to any significant extent, and the blood ammonia concentration rises. In addition, the shunt allows portal blood containing a high concentration of ammonia (see text) to mix directly with the systemic circulation, thus raising the overall blood ammonia level still further. This latter is an invariant feature of portal–systemic shunting, and produces a basal level of blood ammonia which is significantly higher than that found in normal individuals. In some cases, this by itself may be sufficient to produce PSE; in others, PSE is generated by the additional load of exogenously supplied protein. The inability of the liver to handle a nitrogen challenge is known as "protein intolerance" or "ammonia intolerance" and can be generated by, for example, protein feeding or gastrointestinal bleeding. During an episode of PSE, blood ammonia levels can rise as high as 500 μg/ml.

It is important to keep in mind that it is the development of the collateral circulation in response to the cirrhosis, and not the cirrhosis itself, which is responsible for the hyperammonemia. Alcoholism, of course, is a frequent cause of cirrhosis, but it can also occur in the wake of hepatitis, as a result of biliary obstruction, and other causes. In this connection, it should be noted that individuals who have received portocaval shunts can also develop PSE.

INHERITED HYPERAMMONEMIA

PSE, then, is basically a disease of adults with cirrhosis. On the other hand, the hyperammonemia, protein intolerance, and encephalopathy caused by inherited deficiencies of urea cycle enzymes (Table 11.1) are almost always seen only in children without cirrhosis. Other inborn errors of amino acid metabolism such as hyperlysinemia, hyperornithinemia, and others are accompanied by hyperammonemia, and those patients exhibit the characteristic encephalopathy and mental retardation. These will be discussed more fully in the Clin. Corr. for Chapter 12. In this chapter we will consider only deficiencies of urea cycle enzymes.

In general, the severity of the hyperammonemia depends to a large extent on how close the defect is to the point of entry of ammonia into the cycle, and correlates roughly to the severity of the symptoms. Short of death, symptoms can range in seriousness from severe recurrent encephalopathy to transient episodes after a large protein meal, to no symptoms at all, depending roughly upon the degree of enzyme deficiency. However, total or near-total loss of an enzyme is usually fatal to the newborn.

Before continuing the discussion of the deficiencies of urea cycle enzymes, a few general remarks are in order. First, it is important to note that, with the exceptions of a few classic disease states, such as phenylketonuria (to be discussed in Chapter 12), awareness of and interest in congenital defects in amino acid metabolism are relatively recent. As a result, for many of these inherited defects, there are very few case reports. The present paucity of reports is probably more indicative of lack of recognition than of intrinsic rarity of occurrence. Thus, when estimates of frequency of occurrence of a specific defect are given, they should in most cases be taken as minimum estimates.

Second, most of the hereditary defects in amino acid metabolism, including deficiencies of the urea cycle enzymes, produce mental retardation of varying degrees, especially when untreated. This seems to be almost a hallmark of congeni-

tal defects in amino acid metabolism. (This is not to exclude the occurrence of mental retardation with other types of metabolic disorders, only that it seems to appear more often with defects in amino acid metabolism.)

Third, inherited metabolic disorders have an importance far beyond their numbers, because these disorders yield a great deal of information with regard to intermediary metabolism. They are to human biochemistry what bacterial mutants are to

Table 11.1 Deficiencies of Urea Cycle Enzymes

	Mortality and Morbidity							
Enzyme Deficiency	*Total No. of Cases*	*Deaths*	*Mental and/or Physical Retardation*	*Normal*	*Unknown Outcome*	*Blood NH_3*	*Other Urea Cycle Metabolites*	*Other Features*
Carbamoyl phosphate synthetase	6	3	1	2	0	High in 4	Normal ornithine	
Ornithine transcarbamylase	33	17	7	2	7	Elevated; correlates well with symptoms	Normal ornithine	Deficiency demonstrated in Reye's syndrome
Argininosuccinate synthetase	13	7	5	1	0	Variable elevation	Significantly elevated blood citrulline	
Argininosuccinate lyase	42	9	30	3	0	Probably normal	Massively elevated argininosuccinate in blood, CSF, and urine; can detect deficiency by urine screening in absence of symptoms	*Trichorrexis nodosa* ("brittle hair")
Arginase	4	0	4	0	0	Considerably elevated	Arginine significantly elevated in CSF and blood	
Overall totals	98	36	47	8	7			

microbial metabolism. These patients are, in effect, walking laboratories of metabolic pathways.

Finally, a brief general discussion of the detection of these disorders is appropriate at this point. The detection of an inborn error of metabolism begins either from positive findings obtained from routine screening of infants or from observation of behavioral or developmental problems as noted by parent or pediatrician, or both. Often, the key observation is a characteristic odor (see, for example, Clin. Corr. 12.10). These initial observations are usually followed by an analysis of blood and urine for metabolite levels or patterns. This analysis can be carried out by paper chromatography of urine, amino acid analysis of blood by ion-exchange chromatography, and, in some cases, the use of gas–liquid chromatography or mass spectrometry for studying unknown metabolites. [In Hartnup disease (Clin. Corr. 12.11), for example, a characteristic pattern of urinary amino acids is observed with two-dimensional paper chromatography.] Such an analysis will, it is hoped, narrow down the possibilities, and further clarification can come from measurement of enzyme levels in blood and/or tissues obtained from biopsy specimens, or from cells grown in culture.

Against this background we can continue the discussion of deficiencies in urea cycle enzymes. As a group, these enzyme deficiencies are highly lethal. As Table 11.1 shows, close to half of the patients died as a direct result of the deficiency; most of the deaths occurred in the first two weeks of life. Nearly all of the survivors exhibited varying degrees of mental and/or physical retardation. Screening programs designed to detect congenital metabolic disorders fail to pick up many of these cases, mainly because those patients are dead before screening begins, which is usually not until the infants are 3 to 6 weeks old.

All the enzyme deficiencies are characterized by hyperammonemia of varying degrees, especially carbamoyl phosphate synthetase and ornithine transcarbamylase deficiencies. In most cases, the level of the metabolite just prior to the block (Figure 11.12) is considerably elevated in body fluids. Differential diagnosis can be accomplished only by measurement of blood levels of various metabolites and, ultimately, by direct assay of enzyme activities in blood cells, from skin culture, or from liver biopsy or autopsy samples.

SOME THOUGHTS ON THE ROLE OF AMMONIA IN THE TOXIC PROCESS

The episodic encephalopathies seen in children with urea cycle enzyme deficiencies are similar in character to those observed in adults with acquired hyperammonemic conditions. The common link is, of course, the hyperammonemia itself. For example, lowering of blood levels of ammonia by restriction of protein intake in both types of conditions usually results in the termination of the encephalopathic episodes. In fact, reduction of blood ammonia concentrations has been about the only effective treatment for these conditions. Thus the circumstantial evidence for the intrinsic involvement of ammonia in the resulting toxic state is very strong.

Nevertheless, no convincing hypothesis explaining the role of ammonia has yet been put forth. Most of those that have been proposed have as a central feature the notion that the increased blood ammonia concentration interferes with cerebral energy metabolism by draining off α-ketoglutarate from the TCA cycle to form glutamate (see Section 11.2), thus interfering with respiration and ATP produc-

tion. On the face of it, this is an attractive hypothesis, and several pieces of evidence support it. First, in experiments in which rats were given serial doses of ammonia over a 6 h period, GDP rose significantly, and GTP fell significantly, relative to normal levels, a finding consistent with a depletion of α-ketoglutarate. Other experiments of a similar nature showed a depletion of α-ketoglutarate. Second, it is known that when the brain is presented with high levels of ammonia via the carotid artery, the jugular (venous) concentration is considerably lower than the carotid (arterial). Furthermore, oxygen consumption by the brain is decreased by about 50% during hepatic coma.

However, these findings are not conclusive. In particular, the uptake of ammonia by the brain has been accounted for by the formation of glutamine from glutamate in the presence of glutamine synthetase; this reaction is the major pathway of ammonia detoxification in the brain. (However, this does not directly account for the decrease in oxygen consumption.) Experiments in which [^{15}N]ammonia was given to cats showed that there was little change in the

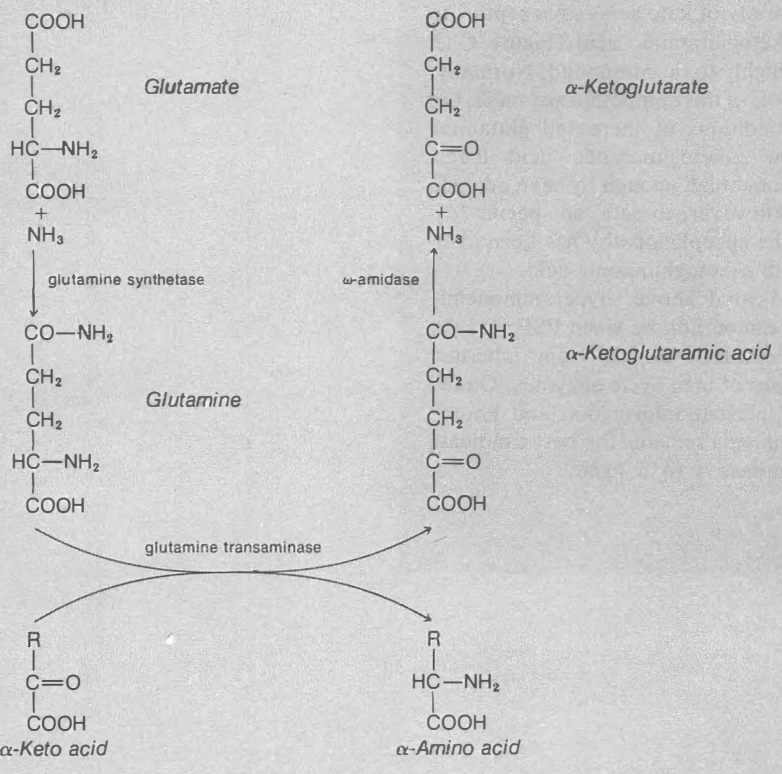


Figure C.C. 11.1
Synthesis and breakdown of α-ketoglutaramic acid.

glutamate concentration and little incorporation of ^{15}N into glutamate, whereas there was a significant increase in the concentration of glutamine and a correspondingly large amount of ^{15}N in amide nitrogen. The results were interpreted as demonstrating the existence of a small but metabolically active pool of glutamate. However, the finding of essentially no change in glutamate would imply a resynthesis of glutamate, presumably at the expense of α-ketoglutarate, consistent with the idea of depletion of TCA cycle intermediates.

Another aspect of the formation of glutamine in brain is the fact that this compound can undergo transamination, with a variety of keto acids as acceptor, to form α-ketoglutaramic acid (Figure C.C. 11.1), a highly toxic compound. Normally, brain levels of this compound are small, but under conditions of increased glutamine formation α-ketoglutaramic acid levels may become high enough to have adverse effects. However, to date, no specific feature of the encephalopathy has been identified with α-ketoglutaramic acid.

As indicated above, hyperammonemia is the common link between PSE and the encephalopathies arising from inherited deficiencies of urea cycle enzymes. On the basis of present information and knowledge, ammonia remains the best candidate for the causative toxic agent.

BIBLIOGRAPHY

See bibliography for Chapter 12.

12

Amino Acid Metabolism II

THOMAS I. DIAMONDSTONE

METABOLISM OF THE INDIVIDUAL AMINO ACIDS

12.1 OVERVIEW

The nutritional requirements of mammals, including man, include certain of the protein amino acids. These amino acids are called "essential" because the animal body cannot synthesize them, but must obtain them preformed in the diet. The others, called "nonessential," can be synthesized from the corresponding α-keto acids, from closely similar compounds, or from essential amino acids. Table 12.1 lists the essential and nonessential amino acids. Arginine is listed as essential because, although the enzyme system exists for synthesizing this amino acid, the presence of large quantities of arginase ensures that most of the arginine formed is cleaved to urea and ornithine (Chapter 11); hence, normal growth and function require more arginine than can be synthesized endogenously. Two of the nonessential amino acids, hydroxyproline and hydroxylysine, are formed from proline and lysine in situ in a polypeptide chain of collagen (Chapter 19).

The metabolism of the individual amino acids presents a picture of considerable complexity, which arises partly from the differences in structure of the various amino acids and partly from the fact that a number of intermediates are utilized for various special functions. Another feature of the complexity has to do with multiple degradative or synthetic pathways, and still another arises because portions of pathways are shared. For the reader unfamiliar with the details, the complexities can be almost overwhelming.

Nevertheless, there are certain patterns and relationships, the un-

Table 12.1 Essential and Nonessential Amino Acids

Essential	*Nonessential*
Arginine	Alanine
Histidine	Asparagine
Isoleucine	Aspartate
Leucine	Cysteine
Lysine	Glutamine
Methionine	Glutamate
Phenylalanine	Glycine
Threonine	Hydroxyproline
Tryptophan	Proline
Valine	Serine
	Tyrosine

derstanding of which will make it easier to deal with the mass of information. During the course of the discussion, these patterns and relationships will be pointed out, and eventually the individual bits and pieces will begin to fall into place, and a certain coherence will develop in regard to the body of data.

The major portion of this chapter deals with the biosynthesis of the nonessential amino acids and the metabolic breakdown of all the important amino acids. We describe the participation of certain amino acids in the biosynthesis of other biologically important compounds and briefly discuss amino acid transport.

12.2 BIOSYNTHESIS OF NONESSENTIAL AMINO ACIDS

For ease of presentation, it is convenient to group the nonessential amino acids as follows: alanine, aspartate, asparagine, glutamate, glutamine, and proline in group 1; and serine, glycine, and cysteine in group 2. The synthesis of tyrosine, which occurs in one step, will be discussed in connection with the metabolism of phenylalanine (see below). The formation of hydroxyproline and hydroxylysine, mentioned above, is discussed in Chapter 19, in the section on the biosynthesis of collagen.

Alanine, Aspartate, Asparagine, Glutamate, Glutamine, and Proline

The simplest biosynthetic pathway is transamination (Chapter 11), and three amino acids of group 1 are regularly synthesized by this route: alanine, aspartate, and glutamate. Either of the first two can donate an amino group to α-ketoglutarate, thus generating glutamate; conversely, glutamate can donate its amino group to either pyruvate or oxalacetate, producing either alanine or aspartate. The reactions involved are catalyzed by either aspartate transaminase (Chapter 11) or *alanine transaminase:*

$$\text{Aspartate} + \alpha\text{-ketoglutarate} \rightleftharpoons \text{oxalacetate} + \text{glutamate}$$

$$\text{Alanine} + \alpha\text{-ketoglutarate} \rightleftharpoons \text{pyruvate} + \text{glutamate}$$

In the reactions as written, glutamate is formed at the expense of

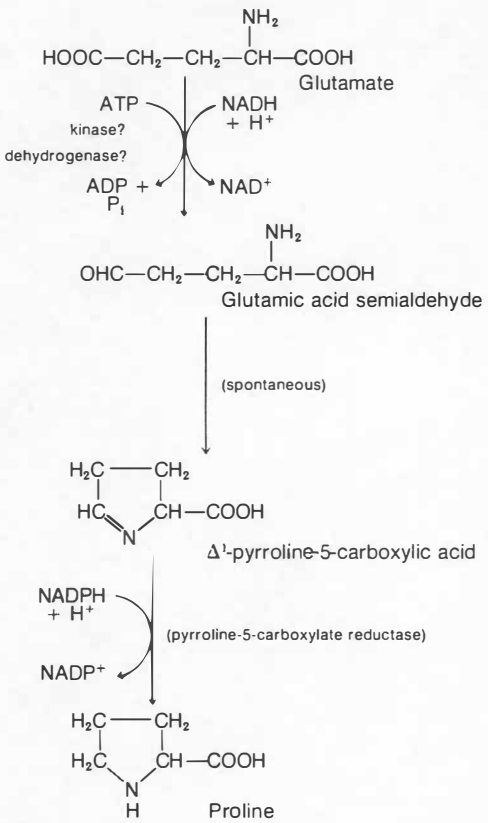


Figure 12.1
The formation of proline from glutamic acid.

aspartate and alanine; in the reverse direction, these two amino acids are formed at the expense of glutamate.

Glutamate can also be formed by the reverse of oxidative deamination, that is, the reductive amination of α-ketoglutarate; the reaction is catalyzed by glutamate dehydrogenase. As discussed in Chapter 11, the formation of glutamate is the thermodynamically favored direction.

An important and widespread reaction is the formation of *glutamine*, catalyzed by *glutamine synthetase*. As indicated in Chapter 11, formation of glutamine is the main mechanism available to the brain for the detoxification of ammonia, and this reaction is also utilized by the liver in removing the ammonia arriving via the portal circulation. The reaction is rather complex, requiring ATP:

$$\text{ATP} + \text{NH}_3 + \text{glutamate} \rightleftharpoons \text{glutamine} + \text{ADP} + \text{P}_i$$

γ-Glutamylphosphate is an enzyme-bound intermediate.

Glutamate is converted to *proline* in a series of reactions involving the formation of *glutamic acid semialdehyde* and Δ^1-pyrroline-5-carboxylic acid (Figure 12.1). The first step requires ATP and NADH, and appears to be analogous to the formation of glyceraldehyde 3-phosphate from 3-phosphoglyceric acid. The formation of the pyrroline carboxylic acid is nonenzymatic; reduction of this compound in the presence of NADPH generates proline.

Asparagine is analogous in structure to glutamine, but the mechanism of its formation in mammalian tissues is not well understood. *Asparagine synthetase* activity has been demonstrated in Novikoff hepatoma and in embryonic chicken liver; both systems require glutamine as amino donor, and the hepatoma enzyme has an explicit requirement for ATP, which is converted to AMP and inorganic pyrophosphate in the course of the reaction. Tumors appear to be capable of much more active synthesis of asparagine than normal tissue.

The synthesis of the six amino acids in group 1 is summarized in Figure 12.2.

Serine, Glycine, and Cysteine

We begin the description of the synthesis of the amino acids in group 2 with *serine*. The major route of formation of this amino acid in mammals is from 3-phosphoglyceric acid. This compound is oxidized by NAD^+ in the presence of *phosphoglycerate dehydrogenase* to yield 3-phosphopyruvate, which undergoes transamination to produce *3-phosphoserine*. This compound is hydrolyzed to free

serine. Alternatively, the 3-phosphoglyceric acid can be converted to 2-phosphoglycerate, which in turn can be dephosphorylated and the resulting glyceric acid oxidized to *hydroxypyruvate;* the latter then undergoes transamination to serine. The series of reactions is shown in Figure 12.3. Glycine and glutamate, as well as alanine, can serve as amino donors to hydroxypyruvate.

Cysteine is synthesized in mammals from methionine and serine in a series of reactions (Figure 12.4), the net effect of which is to replace the hydroxyl oxygen of serine with sulfur. For this reason, the process is often referred to as *transsulfuration.* In the first step, the adenosine moiety of ATP is attached to the sulfur of methionine to form S-*adenosylmethionine;* the reaction is catalyzed by *methionine adenosyltransferase.* This is the only known reaction in which the adenosyl portion of ATP is bound to an acceptor. Besides being an intermediate in this pathway, *S*-adenosylmethionine is important in its own right as a methyl group donor. This function will be discussed in connection with the metabolism of one-carbon units.

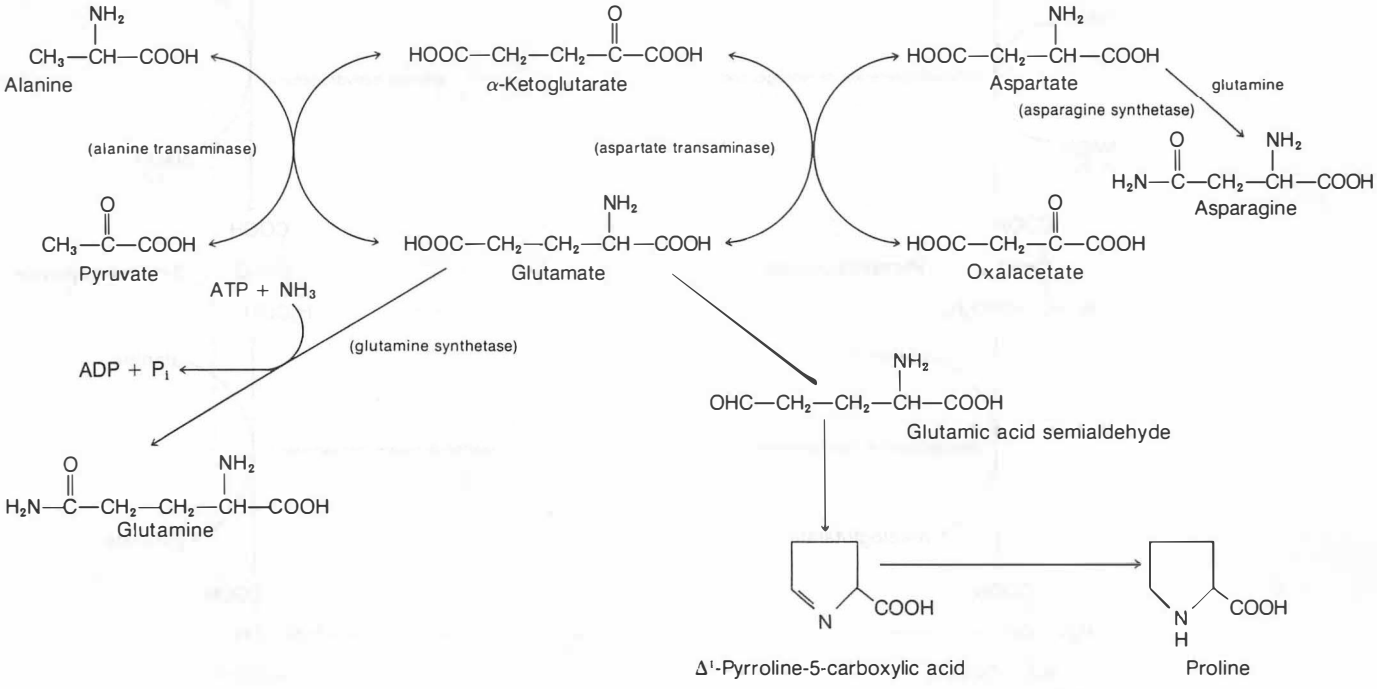


Figure 12.2
Interconversion of alanine, aspartate, and glutamate, and the formation of asparagine, glutamine, and proline.

S-Adenosylmethionine donates its methyl group to an acceptor in the presence of an appropriate methyltransferase to produce a methylated acceptor and S-*adenosylhomocysteine*. The latter compound is cleaved to yield adenosine and *homocysteine*, a reaction catalyzed by *adenosylhomocysteinase*, and the homocysteine condenses with serine in the presence of *cystathionine-β-synthase* to produce *cystathionine*. This compound is cleaved in a reaction catalyzed by *cystathionine-γ-lyase* generating ammonia, *α-ketobutyric acid*, and cysteine. (See Clin. Corr. 12.1.)

In vertebrate livers, the main pathway for the formation of *glycine* is from serine. These two amino acids are interchangeable through

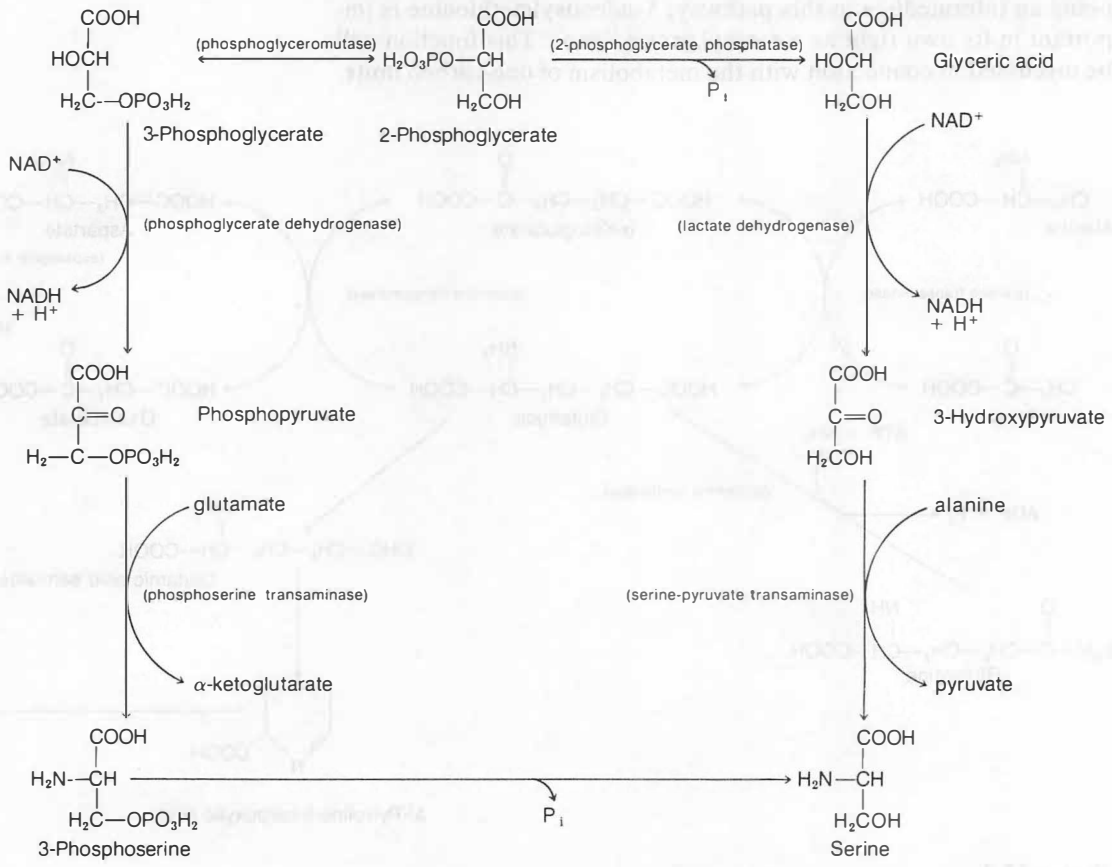


Figure 12.3
Biosynthesis of serine from 3-phosphoglyceric acid.

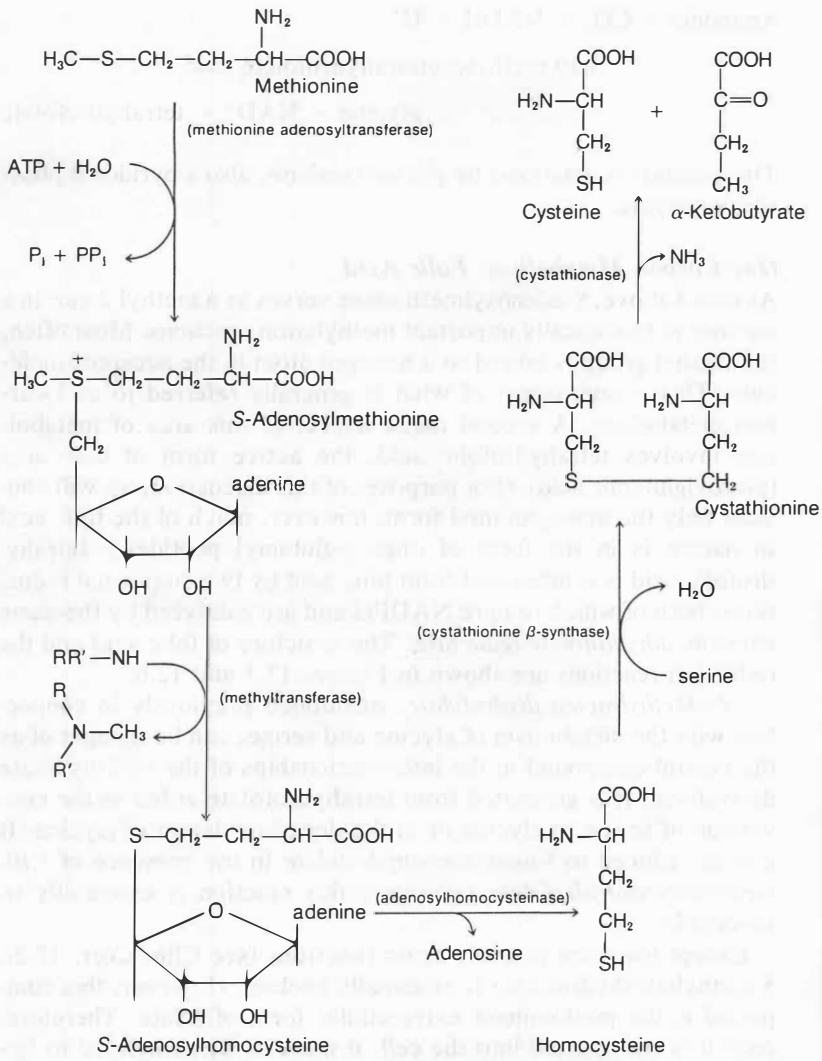


Figure 12.4
Formation of cysteine and α-ketobutyric acid.

the action of *serine hydroxymethyltransferase* with participation of *tetrahydrofolate* (discussed later) and its 5,10-methylene derivative:

Serine + tetrahydrofolate $\rightleftharpoons$
glycine + 5,10-methylenetetrahydrofolate

Serine hydroxymethyltransferase is a pyridoxal phosphate enzyme. Glycine can also be synthesized by a reversal of its decarboxylation:

$$\text{Ammonia} + CO_2 + \text{NADH} + H^+ + \text{5,10-methylenetetrahydrofolate} \rightleftharpoons \text{glycine} + NAD^+ + \text{tetrahydrofolate}$$

This reaction is catalyzed by *glycine synthase*, also a pyridoxal phosphate enzyme.

One-Carbon Metabolism: Folic Acid

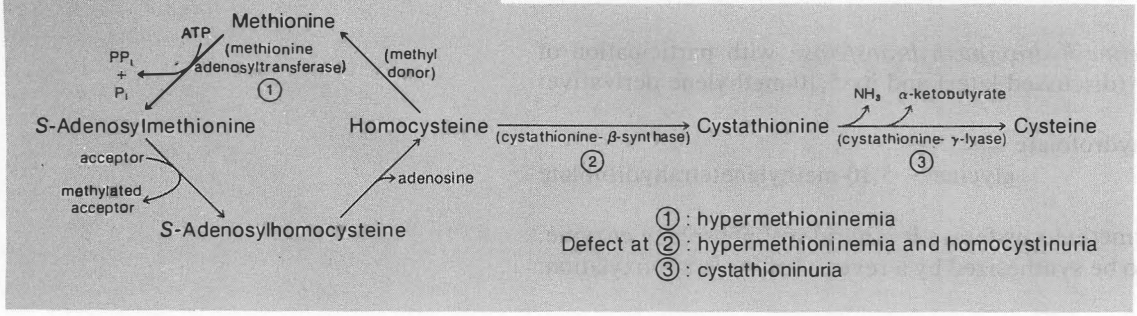
As noted above, *S*-adenosylmethionine serves as a methyl donor in a number of biologically important methylation reactions. Most often, the methyl group is bound to a nitrogen atom in the acceptor molecule. This is one aspect of what is generally referred to as 1-carbon metabolism. A second major aspect of this area of metabolism involves tetrahydrofolic acid, the active form of *folic acid* (pteroylglutamic acid). (For purposes of this discussion, we will consider only the monoglutamyl form; however, much of the folic acid in nature is in the form of oligo-γ-glutamyl peptides.) Tetrahydrofolic acid is synthesized from folic acid by two sequential reductions, both of which require NADPH and are catalyzed by the same enzyme, *dihydrofolate reductase.* The structure of folic acid and the reduction reactions are shown in Figures 12.5 and 12.6.

5,10-Methylenetetrahydrofolate, mentioned previously in connection with the metabolism of glycine and serine, can be thought of as the central compound in the interrelationships of the various folate derivatives. It is generated from tetrahydrofolate either in the conversion of serine to glycine or in the decarboxylation of glycine. It can be reduced to *5-methyltetrahydrofolate* in the presence of *5,10-methylenetetrahydrofolate reductase;* this reaction is essentially irreversible.

Except for some possible brain functions (see Clin. Corr. 12.2), 5-methyltetrahydrofolate is essentially useless. However, this compound is the predominant extracellular form of folate. Therefore, once it is transported into the cell, it must be demethylated to tetrahydrofolate; it cannot be oxidized to 5,10-methylenetetrahydro-

CLIN. CORR. 12.1 DISORDERS OF TRANSSULFURATION

A number of biochemical abnormalities involving sulfur-containing amino acids have been described in recent years. These include hypermethioninemia and hypermethioninuria, homocystinemia and homocystinuria, and cystathioninuria. These abnormalities may be produced by one or more of a number of distinct and specific genetic defects affecting the enzymes involved in the biosynthesis of cysteine from methionine, as shown in the diagram below. To date, three such defects have been described: methionine adenosyltransferase activity, cystathionine-β-synthase deficiency and γ-cystathionase (cystathionine-γ-lyase) deficiency. Of the three, cystathionine-β-synthase deficiency is by far the most common.



METHIONINE ADENOSYLTRANSFERASE DEFICIENCY

Two cases of methionine adenosyltransferase deficiency have been reported, both detected in the course of routine newborn screening programs. In both cases only methionine showed elevated levels, an observation consistent with the location of the block. The most striking clinical finding in these cases was that there were no findings; both children were, and are, perfectly normal and healthy, except for the hypermethioninemia. It has been suggested that normal human liver has a capacity for synthesizing *S*-adenosylmethionine far in excess of need, and that even the greatly reduced activities shown by these patients (12–20% of normal) were still sufficient to meet the requirements of normal health and function.

CYSTATHIONINE-β-SYNTHASE DEFICIENCY

Cystathionine-β-synthase deficiency is characterized by the presence of both hypermethioninemia and homocystinuria, consistent with a defect in the conversion of homocysteine to cystathionine. Presumably the elevated levels of homocysteine lead to an increase in resynthesis of methionine via remethylation. The clinical manifestations frequently include mental retardation, as in the urea cycle enzyme deficiencies. In addition to this, there is usually osteoporosis and a characteristic eye defect known as ectopia lentis, or dislocation of the lens. These are the most commonly observed clinical signs; others include myopia, scoliosis and other skeletal problems and, less frequently, retinal detachment, cataracts, spasticity, and psychiatric disturbances.

Postmortem examinations have revealed the existence of widespread fibrous thickening of the intima of arterial walls, together with increased amounts of collagen in the media. These findings, together with the eye and skeletal problems, have led to the suggestion that all the various manifestations are the result of disturbance of normal collagen metabolism by excess homocysteine, perhaps by interference with cross-linking.

Elevated levels of homocysteine in body fluids without corresponding increases in methionine levels are associated with a failure of remethylation, either from cobalamin deficiency or from a defect in 5,10-methylene-THF reductase, the latter resulting in a lowered level of 5-methyl-THF. This is discussed more fully in Clin. Corr. 12.2.

γ-CYSTATHIONASE (CYSTATHIONINE-γ-LYASE) DEFICIENCY

The first case reported of γ-cystathionase deficiency was a 64-year-old mentally retarded woman who excreted large amounts of cystathionine in her urine. Subsequent searching for additional instances of this enzyme deficiency among the mentally retarded were unsuccessful, suggesting that the connection was accidental. This has turned out to be true; although additional instances of γ-cystathionase deficiency have occurred in patients who are mentally retarded, no clear link has been established between the enzyme defect and any clinical manifestation, including mental retardation. Indeed, the occurrence of clinical problems in patients with γ-cystathionase deficiency is highly variable.

Since this appears to be a "benign" disease, no specific therapeutic measures are necessary. Administration of pyridoxine has, in nearly all cases, reduced the level of cystathionine in the urine. The inci-

dence has been variously estimated at 1 : 18,000 and 1 : 78,000, based on the sensitivity of the screening program used for detection. The mode of inheritance appears to be autosomal recessive.

This defect and methionine adenosyltransferase deficiency may be exceptions to the general observation of association of mental retardation with defects in amino acid metabolism.

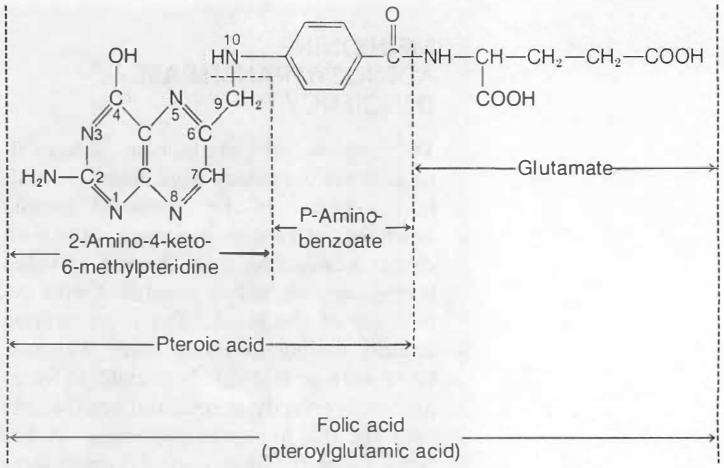


Figure 12.5
Structure of folic acid.

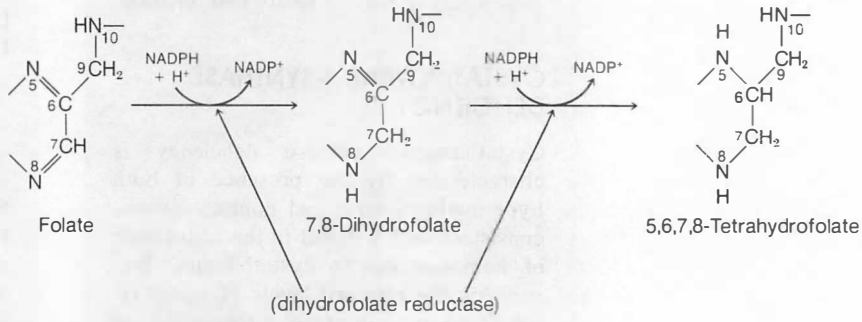


Figure 12.6
Synthesis of dihydrofolate and tetrahydrofolate.

folate because of the irreversibility of the reaction catalyzed by the reductase. The demethylation is accomplished in a reaction in which homocysteine, as the methyl acceptor, is converted back to methionine. The reaction is catalyzed by *5-methyltetrahydrofolate-homocysteine methyltransferase*, and requires the methyl form of vitamin B_{12} (cobalamin) (Figure 12.7). This is the only reaction in which both folate and vitamin B_{12} participate.

This reaction obviously is of central importance in the utilization of folate. However, it is a relatively minor reaction as far as regeneration of methionine is concerned; thus, methionine is an essential

CLIN. CORR. 12.2 NONKETOTIC HYPERGLYCINEMIA AND DISORDERS OF FOLATE METABOLISM

NONKETOTIC HYPERGLYCINEMIA

Nonketotic hyperglycinemia is thought to be caused by a defect in the glycine decarboxylase system, which catalyzes the decarboxylation of glycine to ammonia and carbon dioxide, with the concomitant formation of 5,10-methylene-tetrahydrofolate from tetrahydrofolate (THF). It is characterized by the presence of large quantities of glycine in body fluids unaccompanied by increases in the concentrations of organic acids. This feature distinguishes it from other conditions in which the hyperglycinemia is clearly secondary, such as propionic acidemia and methylmalonic acidemia (Clin. Corr. 12.8).

The plasma concentration of glycine is about twice normal; however, the concentration in cerebrospinal fluid is over 100 times normal, and is much higher in the urine, reaching levels over 1,000 times normal. This is an extremely serious disorder for which there is no treatment at present. Symptoms include severe mental and physical retardation, seizures, overwhelming illness early in life, and spasticity. Many patients do not survive past infancy. The disorder appears to be a rare autosomal recessive trait.

DISORDERS OF FOLATE METABOLISM

Dietary Folate Deficiency

Given the central importance of folate in a variety of metabolic pathways, it is not surprising that deficiency of this nutrient produces a number of symptoms, some well defined, others rather nonspecific. One characteristic is the development of megaloblastic anemia after about 4 months or so of folate deficiency. Only primates and the guinea pig develop this condition, a circumstance which has rendered the use of animal models such as the rat for the study of this condition open to question of relevance. A biochemical manifestation is the excretion of excessive amounts of formiminoglutamic acid (FIGIu) as a result of histidine loading. Severe organic brain damage has been suggested as a possible result of long-term folate deficiency.

Inborn Errors of Folate Metabolism

Congenital Absorption Defect

Four cases of congenital absorption defect have been reported, in which a variety of symptoms was observed; however, common to all these cases was the presence of megaloblastic anemia. Three showed varying degrees of mental retardation. These three patients responded to oral folate therapy; the fourth, who did not suffer from mental retardation, required intramuscular administration. The multiplicity of symptoms and responses suggest the existence of specific carriers both for intestinal transport and across the blood–brain barrier.

Dihydrofolate Reductase Deficiency

Three cases of dihydrofolate deficiency have been reported, all showing the characteristic megaloblastic anemia. The diagnosis was confirmed by liver biopsy, and all three responded only to parenteral 5-formyl-THF, consistent with the nature of the block. (Biologically, any reduced folate would bypass the block, but 5-formyl-THF is the most stable form commercially available, so it is routinely used as a source of reduced folate.) One of

the patients is clinically asymptomatic; the other two are mildly mentally retarded and also show other central nervous system disturbances.

Methylene-THF Reductase Deficiency

Three patients have been described with methylene-THF reductase deficiency, all with central nervous system disturbances of varying nature and severity and also with homocystinuria without hypermethioninemia. Megaloblastic anemia was not mentioned. Two of these patients (sisters) were somewhat mentally retarded; the third was normal in this respect. One of the sisters showed a level of cystathionine-β-synthase activity in liver comparable to that found in heterozygotes for deficiency of that enzyme. Low levels (15–20% of normal) of reductase activity were demonstrated in tissue samples.

The homocystinuria and the central nervous system symptoms do not seem to be connected; however, the latter may arise from diminished amounts of 5-methyl-THF, the level of which is normally higher in cerebrospinal fluid than in serum. 5-methyl-THF participates in a series of reactions in the brain involving tryptamine and some of its derivatives, and it may be that interference with these reactions is connected with the central nervous system symptoms.

THF Methyltransferase Deficiency

The few patients reported to have THF methyltransferase deficiency showed serious failure of central nervous system development, in addition to a variety of other physical findings, including megaloblastic anemia. Most of the patients appear to have disorders of cobalamine metabolism rather than a defect of the enzyme as such. However, low transferase activity was demonstrated in one patient; this was established as a methyltransferase deficiency.

amino acid. The participation of methionine and tetrahydrofolate in 1-carbon transfer is summarized in Figure 12.8.

The mechanism of action of the methyltransferase is not completely understood; however, *S*-adenosylmethionine and a reduced flavin are absolute requirements. In *E. coli*, *S*-adenosylmethionine appears to activate by initially methylating the cobalamin, which then functions as the methyl donor to homocysteine. The cobalamin is then remethylated by 5-methyltetrahydrofolate. The flavin is evidently necessary to maintain the cobalamin in the reduced form. Presumably, the same mechanism operates in mammalian systems. These reactions are summarized in Figure 12.9.

5,10-Methylenetetrahydrofolate is oxidized to *5,10-methenyltetrahydrofolate,* the precursor of *10-formyltetrahydrofolate.* The oxidation is catalyzed by *5,10-methylenetetrahydrofolate dehydrogenase;* the formation of the 10-formyl compound is catalyzed by a *cyclohydrolase.* This compound and its 5,10-precursors play impor-

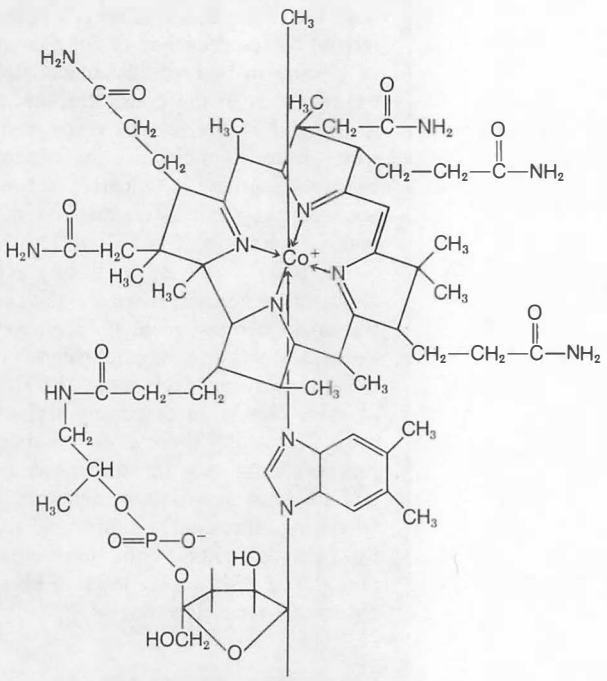


Figure 12.7
Structure of methylcobalamin, the form of vitamin B_{12} that participates in the remethylation of homocysteine to methionine.
Another form, adenosylcobalamin, has an adenosyl residue in place of the methyl group.

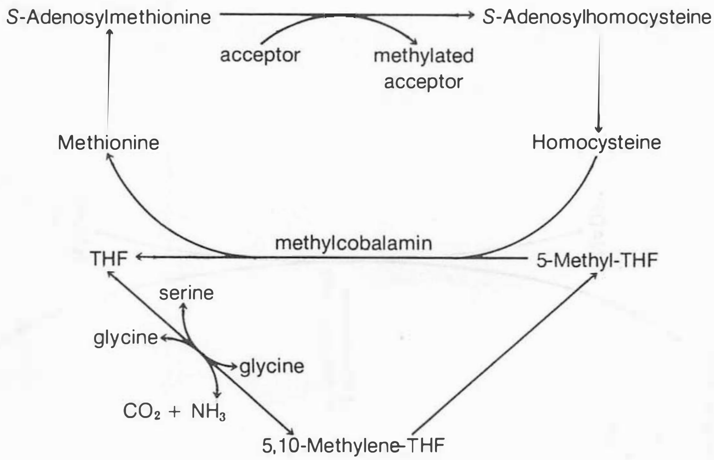
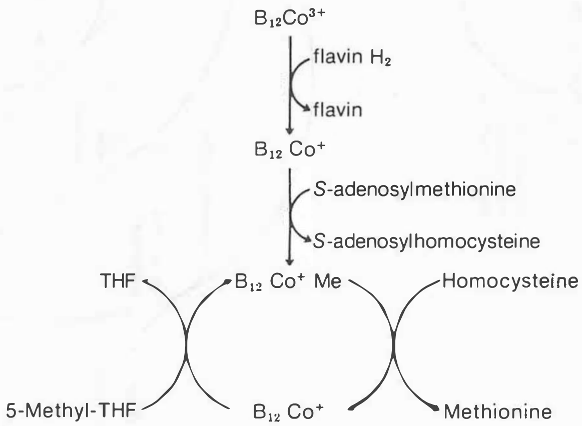


Figure 12.8
Participation of methionine, tetrahydrofolate (THF), and derivatives in 1-carbon metabolism.
See Figure 12.10 for structural details of THF and its derivatives. Note that the lower part of this figure and the upper part of Figure 12.10 are the same.

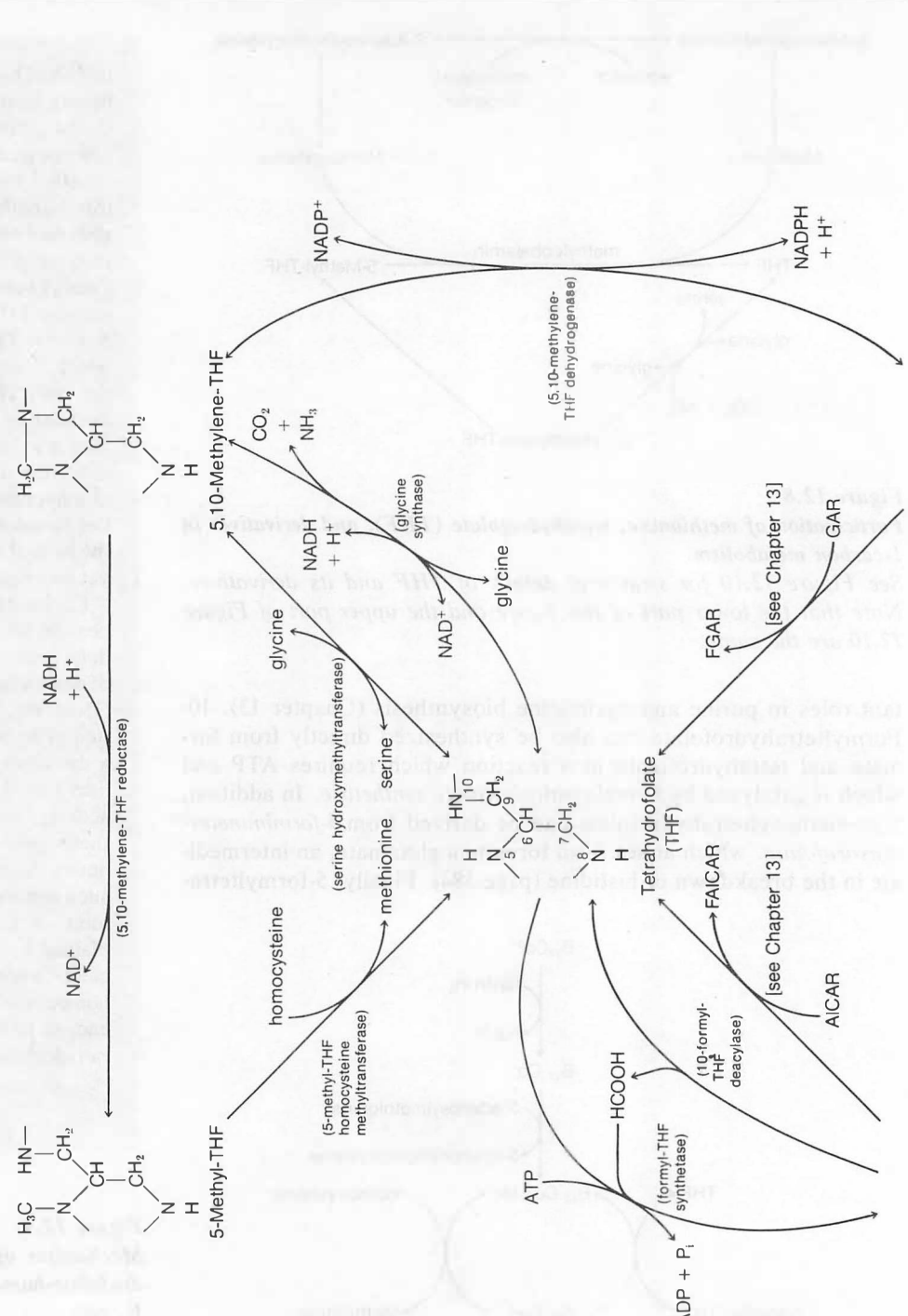
tant roles in purine and pyrimidine biosynthesis (Chapter 13). 10-Formyltetrahydrofolate can also be synthesized directly from formate and tetrahydrofolate in a reaction which requires ATP and which is catalyzed by *formyltetrahydrofolate synthetase.* In addition, 5,10-methenyltetrahydrofolate can be derived from *5-formiminotetrahydrofolate,* which arises from formiminoglutamate, an intermediate in the breakdown of histidine (page 584). Finally, 5-formyltetra-



tablished by enzymatic assay, using a liver biopsy sample. The consequences of reduced activity of this enzyme have been summarized in a concept known as the "methyl-trap" hypothesis. According to this hypothesis, lowered activity of this enzyme (whether from actual deficiency of enzyme protein or from cobalamine deficiency) results in lowering of the level of cellular THF, raising the level of cellular 5-methyl-THF, and reducing the rate at which 5-methyl-THF is transported into the cell. The net effect is to reduce the availability of methyl groups; in effect, they are "trapped" as 5-methyl-THF. The reduction in DNA synthesis predicted by this hypothesis can be invoked to explain the megaloblastic anemic, but at present, the central nervous system degeneration is yet to be accounted for.

Curiously, homocystinuria does not appear to be generally associated with this defect, as might be expected from the reduced availability of methyl groups. (However, see Clin. Corr. 12.8 for a mention of homocystinuria in connection with a disturbance in methylmalonate metabolism.) In the case where low transferase activity was demonstrated (see above), there were no abnormalities in levels of homocysteine and methionine. The lack of such abnormalities, coupled with the symptoms of folate deficiency, might be explained by assuming that there was sufficient transferase activity for normal homocysteine remethylation but not enough to maintain adequate levels of tetrahydrofolate.

Figure 12.9
***Mechanism of action of 5-methylenetetrahydrofolate-homocysteine methyltransferase in* E. coli.**


5-Methyl-THF
NADH + H+
(5,10-methylene-THF reductase)
NAD+
5,10-Methylene-THF
NADP+
NADPH + H+
(5,10-methylene-THF dehydrogenase)
CO2 + NH3
NADH + H+
(glycine synthase)
NAD+
glycine
glycine
(serine hydroxymethyltransferase)
serine
homocysteine
(5-methyl-THF homocysteine methyltransferase)
methionine
Tetrahydrofolate (THF)
FGAR
[see Chapter 13]
GAR
FAICAR
[see Chapter 13]
AICAR
(10-formyl-THF deacylase)
HCOOH
ATP
(formyl-THF synthetase)
ADP + Pi

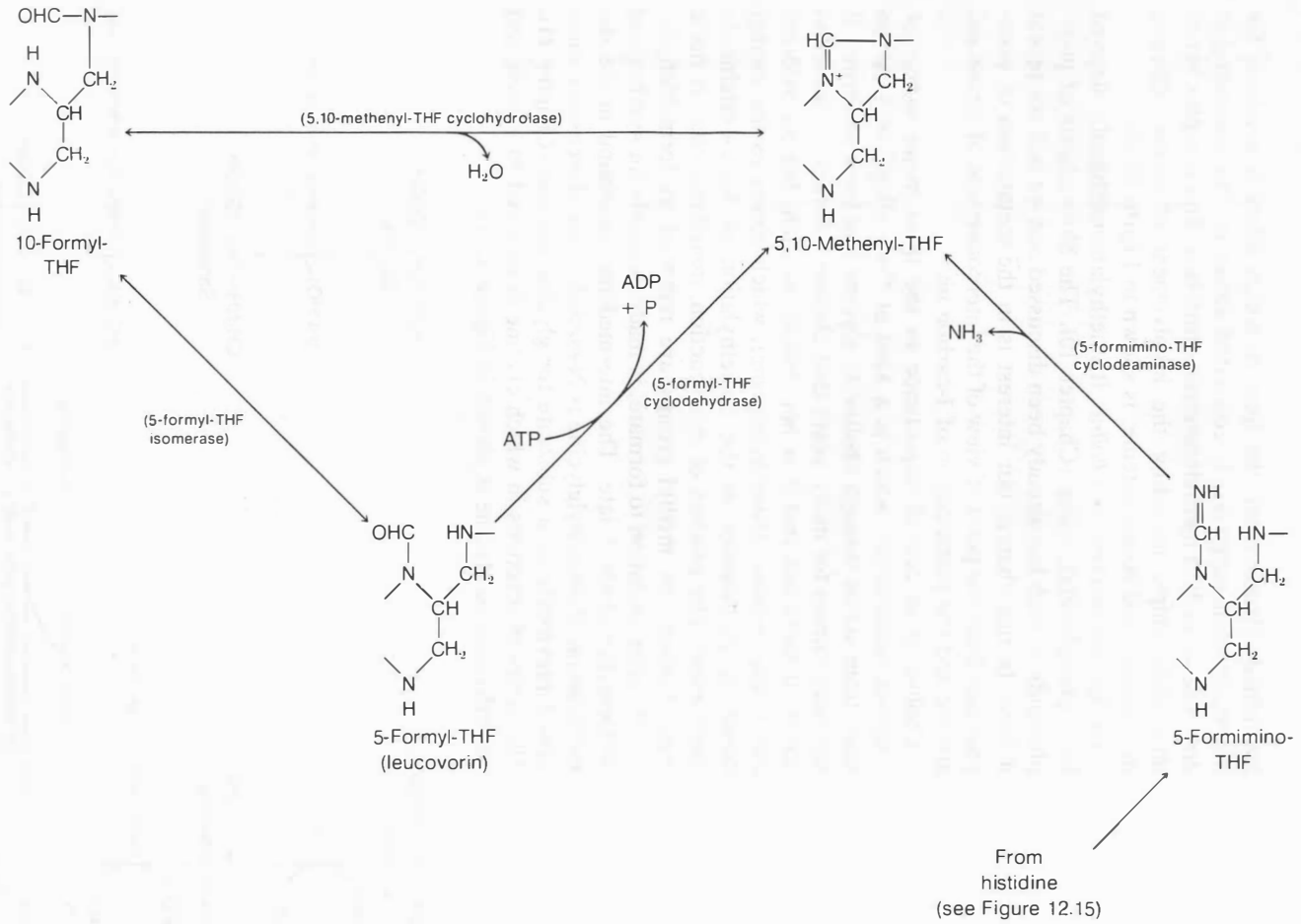


Figure 12.10
One-carbon metabolism, with emphasis on the interconversion of the various folate derivatives and the participation of serine and methionine (cf. Figure 12.8).

hydrofolate (leucovorin), the form in which folate is available for pharmaceutical purposes, is converted either to 10-formyltetrahydrofolate or to 5,10 methylenetetrahydrofolate. The complex set of interrelationships, including the involvement of serine, glycine, methionine, and homocysteine, is shown in Figure 12.10.

Of special interest is *choline* (trimethylaminoethanol), derived from phosphatidylcholine (Chapter 10). The biosynthesis of phospholipids as such has already been discussed and we will not repeat it here. In this chapter our interest is in the metabolism of phospholipids from the point of view of the interconversion of serine and glycine and the participation of 1-carbon units.

Choline is of central importance in the third major aspect of 1-carbon metabolism, which is a kind of "recycling" of 1-carbon units from serine through choline to glycine and back to serine. It has been known for many years that choline functions as a methyl donor. It turns out that it is not choline as such, but an oxidized metabolite, *betaine* (trimethylglycine), which serves as the methyl donor; it participates in the remethylation of homocysteine to methionine. The product of this reaction, *dimethylglycine*, is not a methyl donor; its methyl groups are removed as formaldehyde, which, after oxidation to formate, is made available for synthesis of 10-formyltetrahydrofolate. The intermediate compound in the demethylation of dimethylglycine is N-*methylglycine* (*sarcosine*), mentioned previously as a substrate for glycine oxidase (Chapter 11). The series of reactions in which choline is oxidized to betaine and demethylated to glycine is shown in Figure 12.11.

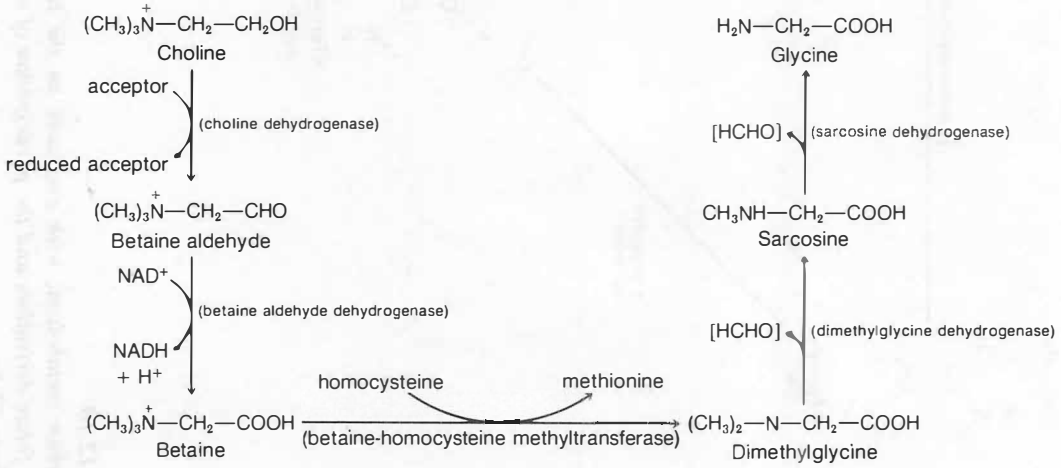


Figure 12.11
Conversion of choline to glycine.

The electron acceptor in the oxidation of choline to betaine aldehyde is unknown; it does not appear to be a pyridine nucleotide. Similarly, the electron acceptors in the oxidation of dimethylglycine and sarcosine are also unknown; however, it is known that the enzymes catalyzing the latter two reactions are flavoproteins. Also, there is evidence that electron transfer from dimethylglycine and sarcosine to the electron transport chain is mediated by the same electron transfer protein as the one that serves this function for fatty acid oxidation.

Obviously, the pathway through choline is not simply an elaborate way of obtaining glycine from serine. Probably the major significance relates to the biosynthesis of phospholipids; however, the conversion of choline to glycine is accomplished in such a way as to make it possible to recover the methyl groups. For this reason, the pathway from choline to glycine can be thought of as a "salvage" pathway for 1-carbon units.

It should be pointed out that in the conversion of serine to glycine

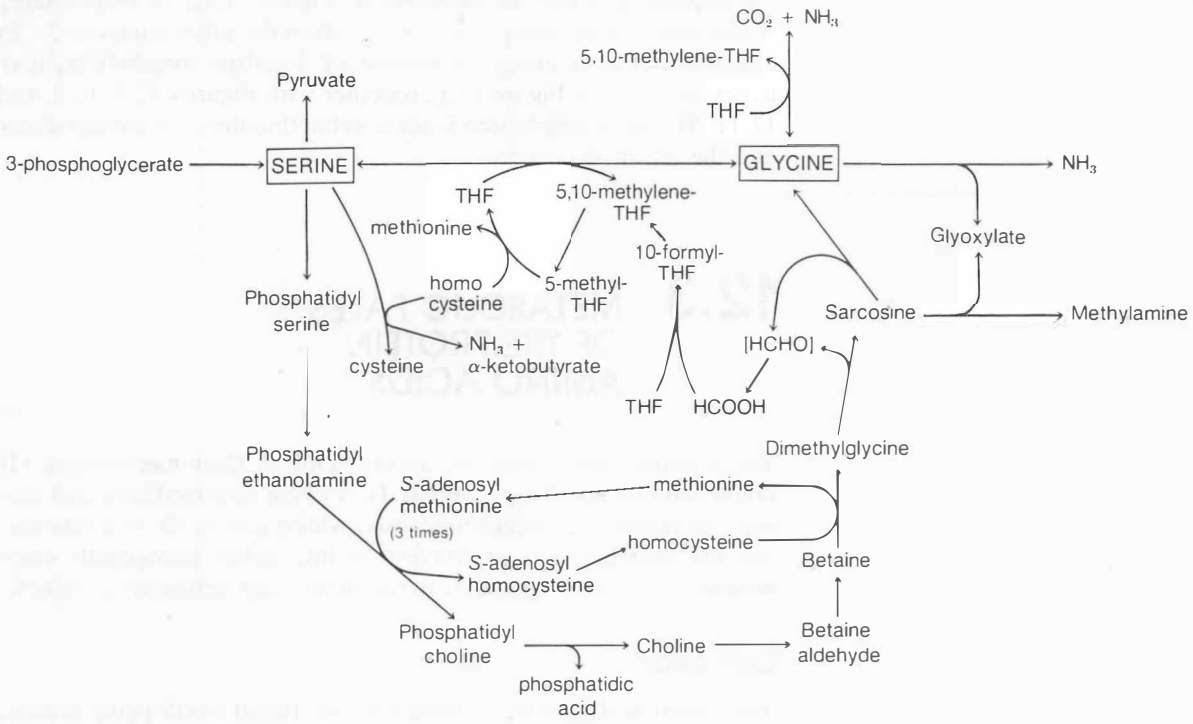


Figure 12.12
The 1-carbon "salvage" pathway involving serine, choline, and glycine.
Note the remethylation of homocysteine by betaine.

Table 12.2 Glycogenic and Ketogenic Amino Acids

Glycogenic	*Glycogenic and Ketogenic*	*Ketogenic*
Alanine	Isoleucine	Leucine
Arginine	Phenylalanine	Lysine
Asparagine	Threonine	
Aspartate	Tryptophan	
Cysteine	Tyrosine	
Glutamine		
Glutamate		
Glycine		
Histidine		
Hydroxyproline		
Isoleucine		
Methionine		
Phenylalanine		
Proline		
Serine		
Threonine		
Tryptophan		
Tyrosine		
Valine		

via phospholipids, the other aspects of 1-carbon metabolism appear, at least in part. *S*-Adenosylmethionine is involved in the synthesis of phosphatidylcholine, and betaine supplies methyl groups for the formation of methionine from homocysteine. Furthermore, as noted above, the methyl groups released from dimethylglycine are available for incorporation into folate derivatives. This is also the fate of the first methyl group released from betaine, if no homocysteine is available as acceptor.

A summary of the relationships involved in 1-carbon metabolism is shown in Figure 12.12. It should be noted that although certain metabolites (methionine, homocysteine, tetrahydrofolate) appear more than once, this does not imply any compartmentalization; the repetition is merely for convenience in display. Although cysteine is not directly involved in 1-carbon metabolism, it is included in this group because it is formed from serine and methionine, which are directly involved. These two amino acids, together with glycine, homocysteine, and the tetrahydrofolate derivatives, are the major participants in 1-carbon metabolism. Figure 12.12, in emphasizing serine and glycine, has given short shrift to the other compounds. In order to obtain a complete picture of 1-carbon metabolism, it is necessary to view Figure 12.12 together with Figures 12.4, 12.8, and 12.11. The latter emphasize *S*-adenosylmethionine as a methyl donor and the tetrahydrofolates.

12.3 METABOLIC FATES OF THE PROTEIN AMINO ACIDS

Three major fates await the amino acids in their metabolism: (1) conversion to acetyl CoA and/or TCA cycle intermediates and ultimate oxidation; (2) gluconeogenesis, which also leads to oxidation; and (3) incorporation or conversion into other nitrogenous compounds, particularly purines, pyrimidines, and neurogenic amines.

Oxidation

The amino acids can be divided into two broad overlapping groups, depending upon the nature of their ultimate oxidation products; they are either glycogenic or ketogenic (Table 12.2). The former yields pyruvate and/or TCA cycle intermediates, and thus provides substrates for gluconeogenesis (page 607); the latter yields one or more

of acetyl CoA, acetoacetyl CoA, or free acetoacetate. The last compound is one of the classic triad of so-called "ketone bodies," the others being β-hydroxybutyrate and acetone. Leucine and lysine are the only amino acids that are not glycogenic.

Figure 12.13 shows a flow chart of amino acid oxidation arranged to show the amino acids grouped according to the glycogenic and/or ketogenic end product. This arrangement also permits the identification of common intermediates for three pairs of amino acids: tryptophan and lysine proceeding through α-ketoadipic acid, threonine and methionine through α-ketobutyrate, and isoleucine and valine through propionyl CoA. It should also be noted that two groups, the oxalacetate group and the α-ketoglutarate group, contain exclusively glycogenic amino acids, and that the amino acids in these two

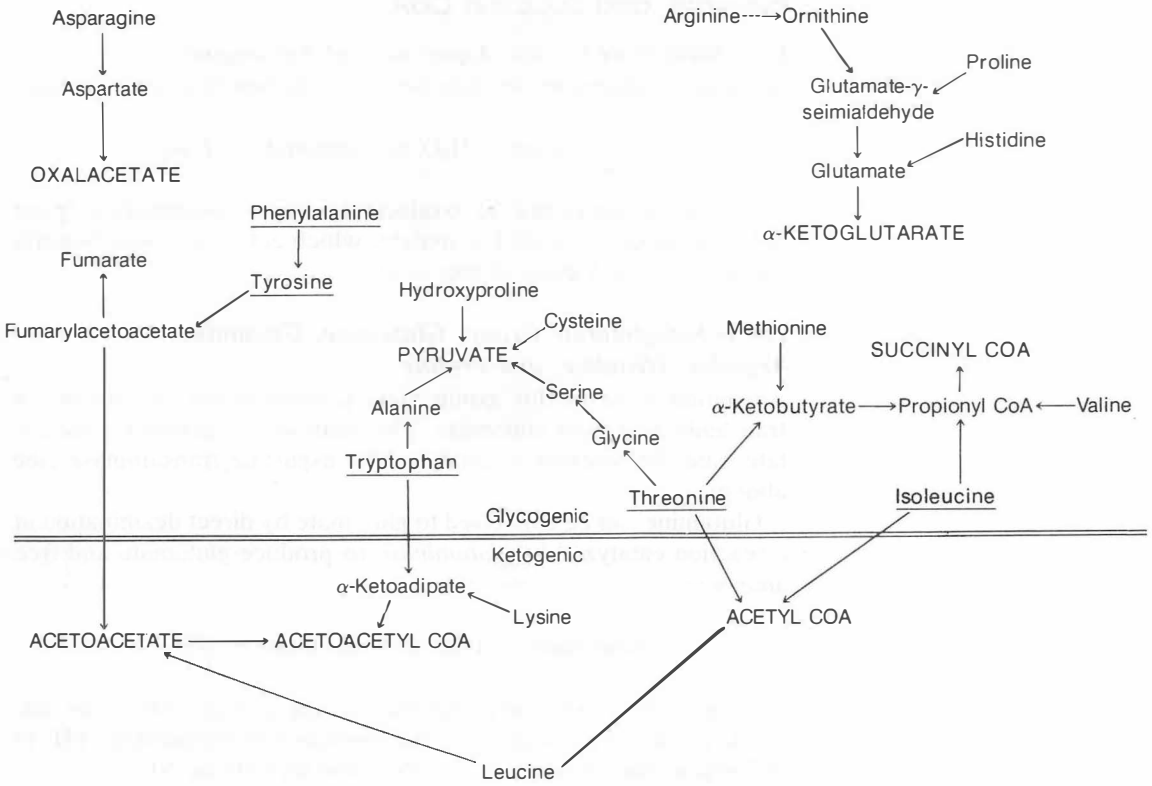


Figure 12.13
Flow chart of oxidative breakdown of the protein amino acids, grouped by end product.
Amino acids underlined are both glycogenic and ketogenic.

groups are not part of any other group, that is, they yield a single end product.

Hydroxylysine does not appear here because once formed, it is not degraded. It is excreted in the urine either intact or as an *O*-glycosylated compound.

In the discussion of the oxidation of the individual amino acids, the groups will be taken in the following order: oxalacetate, α-ketoglutarate, fumarate, pyruvate, succinyl CoA, acetyl CoA, and acetoacetate/acetoacetyl CoA. The reader will find it useful to refer to Figure 12.13 in the course of studying the breakdown of the individual amino acids.

Glycogenic Amino Acids: End Products Oxalacetate, α-Ketoglutarate, Fumarate, Pyruvate, and Succinyl CoA

The Oxalacetate Group: Aspartate and Asparagine

Asparagine undergoes deamination in the presence of *asparaginase:*

$$\text{Asparagine} + H_2O \rightleftharpoons \text{aspartate} + NH_3$$

Aspartate is converted to oxalacetate via transamination (page 565); the latter is reduced to malate, which enters the mitochondria and into the TCA cycle (Chapter 6).

The α-Ketoglutarate Group: Glutamate, Glutamine, Arginine, Histidine, and Proline

All amino acids in this group yield α-ketoglutarate by means of transamination from glutamate. The main keto acceptor is oxalacetate, and the reaction is catalyzed by aspartate transaminase (see above).

Glutamine can be converted to glutamate by direct deamination in a reaction catalyzed by *glutaminase* to produce glutamate and free ammonia:

$$\text{Glutamine} + H_2O \rightleftharpoons \text{glutamate} + NH_3$$

This reaction is especially important in the kidney, where the ammonia produced participates in the regulation of extracellular pH. In this organ, the ammonia enters the urine directly as NH_4^+.

The first step in the catabolism of arginine, cleavage to urea and ornithine, has been discussed in connection with the urea cycle (Chapter 11). Ornithine undergoes transamination to *glutamic acid γ-semialdehyde;* this reaction is catalyzed by *ornithine transaminase. Proline* is oxidized to Δ^1-pyrroline-5-carboxylic acid by the action of

proline oxidase. (Note that in the synthesis of proline from glutamate the reverse of this reaction is catalyzed by pyrroline-5-carboxylate reductase.) The ring compound is in equilibrium with its open chain form, glutamic acid γ-semialdehyde; this equilibration takes place nonenzymatically. Glutamate is generated from the aldehyde by oxidation; the reaction is catalyzed by *1-pyrroline dehydrogenase*. This enzyme also catalyzes the oxidation of 1-pyrroline-3-hydroxy-5-carboxylic acid (from hydroxyproline) to γ-erythro-*hydroxyglutamate* (Figure 12.20). The catabolism of arginine and proline to glutamate is shown in Figure 12.14. The conversion of ornithine to glutamate provides another interconnection between the urea cycle and the TCA cycle.

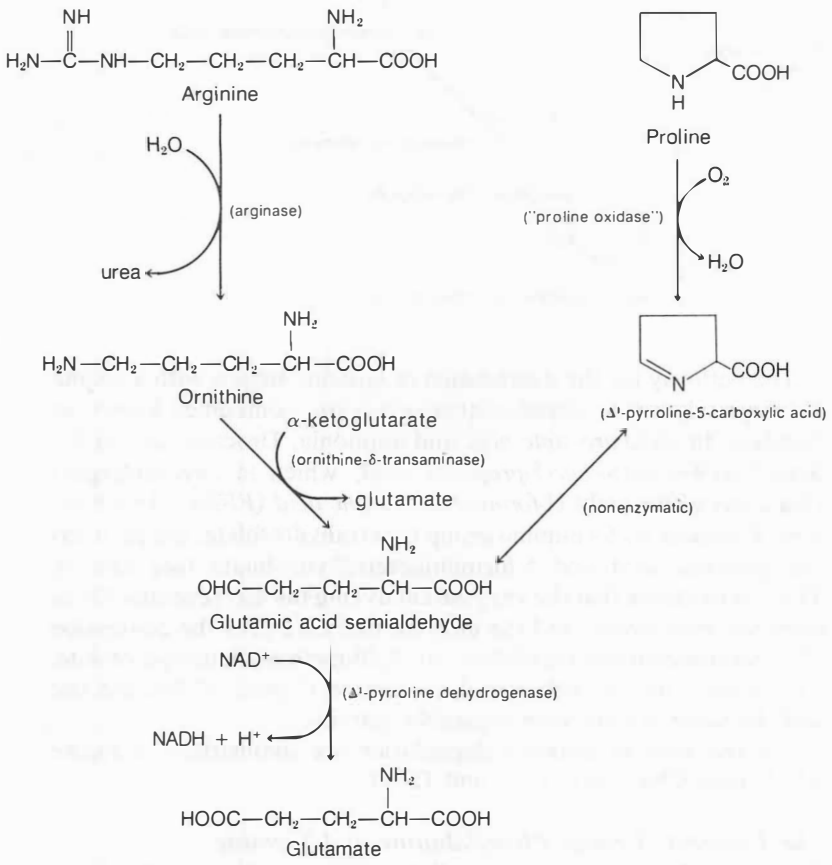


Figure 12.14
The conversion of arginine and proline to glutamate.

CLIN. CORR. 12.3 HISTIDINEMIA

Histidinemia is characterized by elevated levels of histidine in blood. There is also an increased excretion of histidine in the urine, but this finding does not appear to be as specific an indicator for the disorder as blood histidine. Mental retardation and speech impairment are common findings; however, the large variability in incidence of these findings and the significant number of histidinemic patients of normal intelligence precludes making definite correlations. Nevertheless, most of the patients had some sort of developmental defect, in some cases no more than subnormal height and weight. The reported incidence is about 1 in 15,000.

In addition to the high levels of blood histidine, significant quantities of imidazolepyruvic acid were observed in urine. This compound is an aryl-substituted pyruvic acid resembling phenylpyruvic acid, and the urine of affected individuals gave the characteristic green color with ferric chloride, as seen in phenylketonuria. These two findings—elevated blood histidine and elevated urinary imidazolepyruvic acid—point to a defect in histidine ammonia-lyase (Figure 12.15), which was confirmed by direct enzymatic assay in skin samples. The product of the reaction catalyzed by this enzyme is urocanic acid, a compound normally present in sweat; it is absent in histidinemic patients. The imidazolepyruvic acid arises from transamination of histidine, the first step in a relatively minor pathway of histidine degradation.

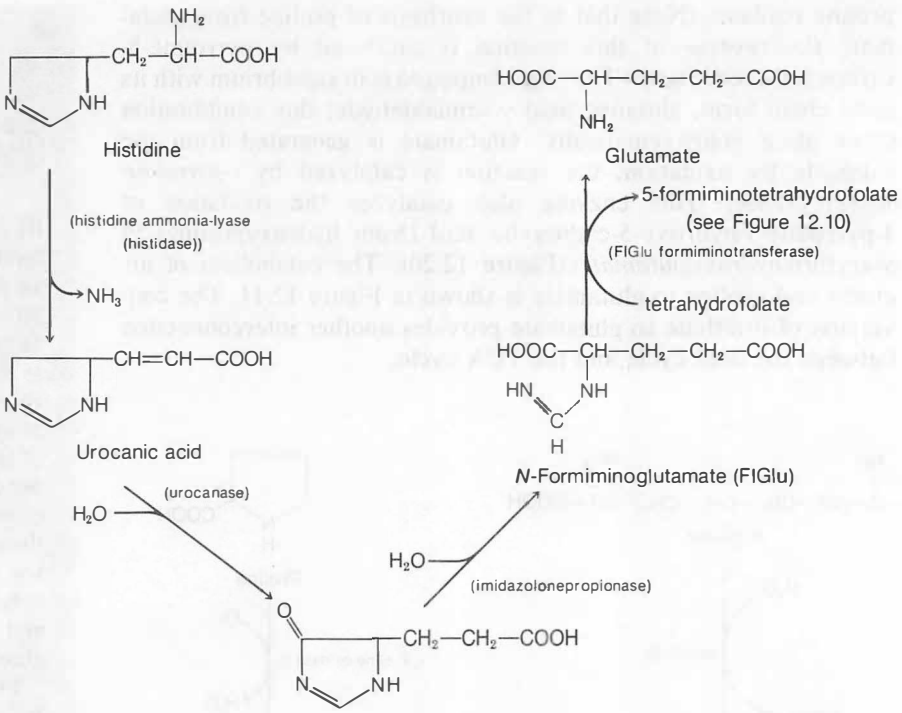


Figure 12.15
The metabolism of histidine.

The pathway for the degradation of *histidine* begins with a deamination catalyzed by *histidine ammonia-lyase,* sometimes known as *histidase,* to yield *urocanic acid* and ammonia. Urocanic acid is hydrated to *4-imidazolone-5-propionic acid,* which in turn undergoes ring cleavage to yield N-*formiminoglutamic acid* (*FIGlu*). This compound donates its formimino group to tetrahydrofolate; the products are glutamic acid and 5-formiminotetrahydrofolate (see above). There is evidence that the enzyme catalyzing the last reaction, *FIGlu formiminotransferase,* and the enzyme that catalyzes the conversion of 5-formiminotetrahydrofolate to 5,10-methenyltetrahydrofolate, *5-formiminotetrahydrofolate cyclodeaminase* (Figure 12.10), are one and the same protein with separable activities.

The reactions of histidine degradation are summarized in Figure 12.15. (See Clin. Corr. 12.3 and 12.4.)

CLIN. CORR. 12.4 FORMIMINOTRANSFERASE DEFICIENCY SYNDROMES

The formiminotransferase deficiency syndromes are characterized by excretion of large amounts of FIGlu (Figure 12.15)—in

The Fumarate Group: Phenylalanine and Tyrosine

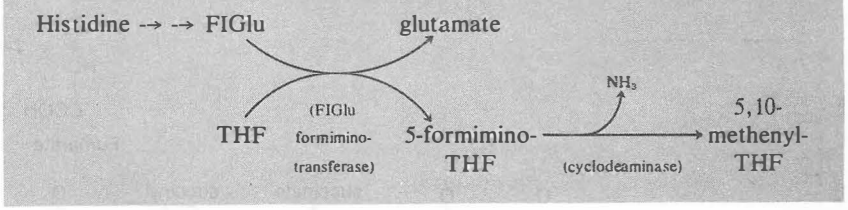
It is somewhat pretentious to call this a "group," since the direct and only precursor of *tyrosine* is *phenylalanine,* and, aside from incorporation into protein, phenylalanine has no other metabolic role

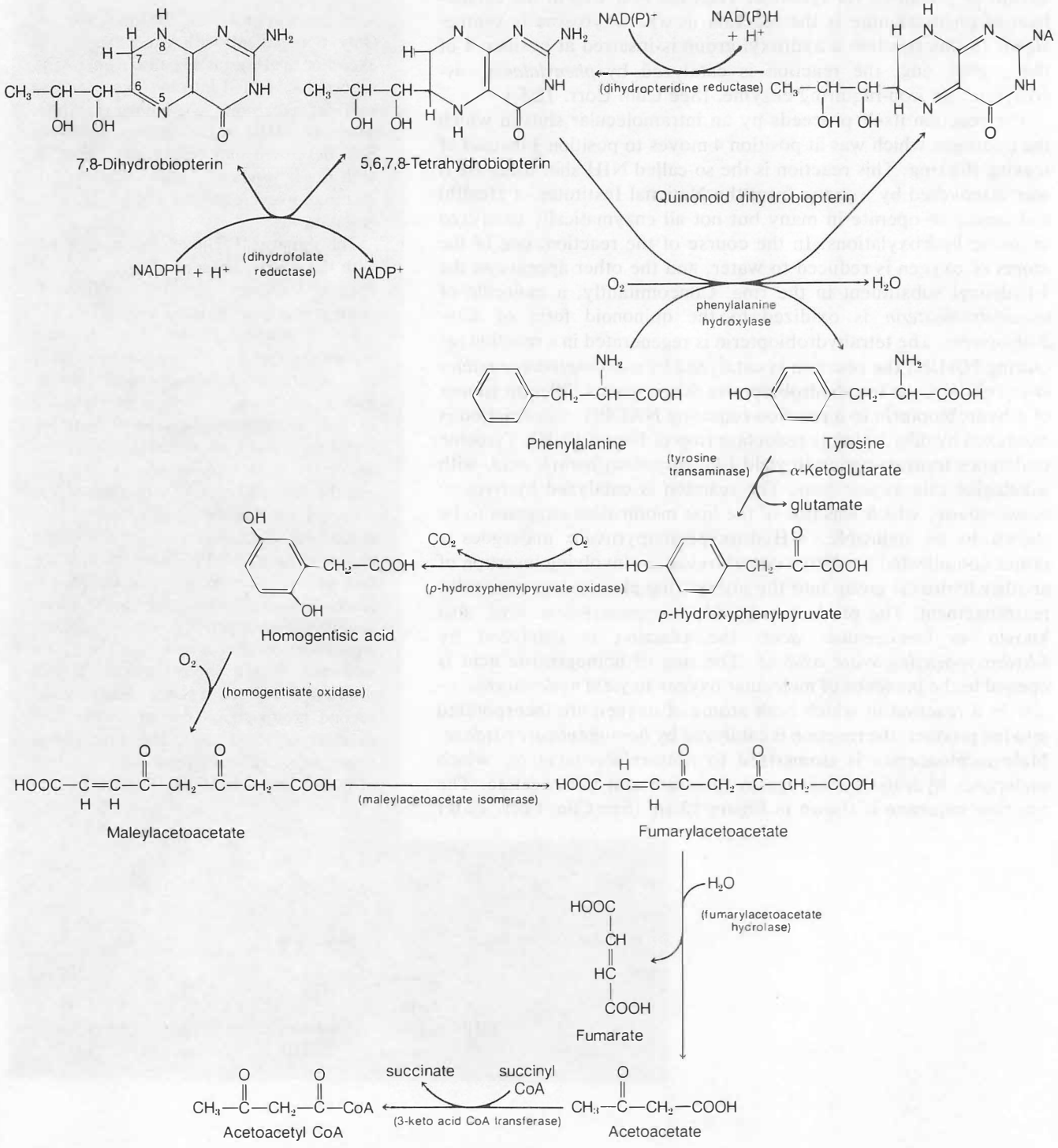
except as precursor for tyrosine. Thus the first step in the catabolism of phenylalanine is the reaction in which tyrosine is synthesized. In this reaction a hydroxyl group is inserted at position 4 of the phenyl ring; the reaction is catalyzed by *phenylalanine hydroxylase,* an iron-requiring enzyme. (See Clin. Corr. 12.5.)

The reaction itself proceeds by an intramolecular shift in which the hydrogen which was at position 4 moves to position 3 instead of leaving the ring. This reaction is the so-called NIH shift (because it was discovered by a group from the National Institutes of Health) and seems to operate in many but not all enzymatically catalyzed aromatic hydroxylations. In the course of the reaction, one of the atoms of oxygen is reduced to water, and the other appears as the 4-hydroxyl substituent in the ring. Concomitantly, a molecule of *tetrahydrobiopterin* is oxidized to the quinonoid form of *dihydrobiopterin.* The tetrahydrobiopterin is regenerated in a reaction requiring NADH; the reaction is catalyzed by *dihydropteridine reductase.* Initially, the tetrahydrobiopterin arises from a different isomer of dihydrobiopterin in a reaction requiring NADPH; this reaction is catalyzed by dihydrofolate reductase (top of Figure 12.16). Tyrosine undergoes transamination to yield *4-hydroxyphenylpyruvic acid,* with α-ketoglutarate as acceptor. The reaction is catalyzed by *tyrosine transaminase,* which was one of the first mammalian enzymes to be shown to be inducible. 4-Hydroxyphenylpyruvate undergoes a rather complicated oxidative decarboxylation involving insertion of another hydroxyl group into the phenyl ring plus an intramolecular rearrangement. The product is *2,5-dihydroxyphenylacetic acid,* also known as *homogentisic acid;* the reaction is catalyzed by *4-hydroxyphenylpyruvate oxidase.* The ring of homogentisic acid is opened in the presence of molecular oxygen to yield *maleylacetoacetate* in a reaction in which both atoms of oxygen are incorporated into the product; the reaction is catalyzed by *homogentisate oxidase.* Maleylacetoacetate is isomerized to *fumarylacetoacetate,* which undergoes hydrolytic cleavage to fumarate and acetoacetate. The reaction sequence is shown in Figure 12.16. (See Clin. Corr. 12.6.)

some cases approaching 100 times normal. Only one patient with excessive FIGlu excretion had megaloblastic anemia. The reported cases fall into two more or less well-defined groups, one having very high levels of FIGlu excretion and relatively mild developmental difficulties, the other with lower levels of FIGlu excretion but with more severe mental and physical retardation.

The abnormal FIGlu excretion, together with the virtual absence of megaloblastic anemia, indicates that this condition is more properly considered a defect in histidine metabolism, rather than of folate metabolism per se. A proposal to explain the difference in biochemical abnormalities has been put forth, which is based on the assumption that the formiminotransferase and cyclodeaminase activities are on the same enzyme protein (see text and the scheme below). According to this proposal, the postulation of a severe block in the formiminotransferase activity would explain the massive FIGlu excretion in the first group, whereas a block in the cyclodeaminase activity, leaving the formiminotransferase activity relatively intact, would account for the relatively low FIGlu excretion of the second group. If this proposal is correct, it is not clear why the second group should be so much more severely affected than the first, since either defect effectively prevents recovery of the 1-carbon unit from histidine.




7,8-Dihydrobiopterin
5,6,7,8-Tetrahydrobiopterin
NAD(P)⁺
NAD(P)H + H⁺
(dihydropteridine reductase)
Quinonoid dihydrobiopterin
NADPH + H⁺
(dihydrofolate reductase)
NADP⁺
O₂
H₂O
phenylalanine hydroxylase
Phenylalanine
Tyrosine
(tyrosine transaminase)
α-Ketoglutarate
glutamate
Homogentisic acid
CO₂
O₂
(p-hydroxyphenylpyruvate oxidase)
p-Hydroxyphenylpyruvate
(homogentisate oxidase)
Maleylacetoacetate
(maleylacetoacetate isomerase)
Fumarylacetoacetate
H₂O
(fumarylacetoacetate hydrolase)
Fumarate
succinate
succinyl CoA
(3-keto acid CoA transferase)
Acetoacetyl CoA
Acetoacetate

The Pyruvate Group: Alanine, Threonine, Serine, Glycine, Cysteine, Tryptophan, and Hydroxyproline

The transamination of alanine to pyruvate has already been discussed. Threonine, serine, and glycine can be considered together. Threonine is cleaved to glycine and acetaldehyde in the presence of serine hydroxymethyltransferase, and it can also be deaminated to α-ketobutyric acid by the action of *threonine dehydratase.* The glycine is converted to serine (see above) which in turn is dehydrated and deaminated in the same reaction to yield pyruvate, analogous to the deamination of threonine; the reaction is catalyzed by *serine dehydratase.* Alternative routes for the disposal of glycine and serine exist; glycine can be decarboxylated to CO_2 and ammonia, the usual direction for the reaction on page 570, and the participation of serine in the formation of phospholipids has already been described. The various interconversions are shown in Figure 12.17.

In mammals, *cysteine* is degraded to pyruvate, as shown in Figure 12.18. The step from *cysteinesulfinic acid* to pyruvate is catalyzed by aspartate transaminase; the immediate product, *β-sulfinylpyruvate,* is unstable and breaks down with release of the sulfite ion. In some mammalian systems, cysteinesulfinic acid is the precursor of *taurine* via either hypotaurine or cysteic acid, as shown. In man, however, this pathway does not appear to operate, at least to any great extent, and the bulk of the taurine is derived either from cysteamine or from dietary sources. Taurine is concentrated to a large extent in heart muscle; the isethionic acid found there is derived from taurine.

The conversion of *β-mercaptopyruvate* to pyruvate can be carried out with any of four different sulfur acceptors, all of which can be utilized by *β-mercaptopyruvate transsulfurase:* cysteinesulfinic acid, sulfite, cyanide, or a sulfhydryl compound (e.g., glutathione). The thiosulfate produced when sulfite is acceptor can be oxidized to sulfate in the presence of cyanide; this reaction is catalyzed by *rhodanese.*

Sulfite can be excreted unchanged or oxidized to sulfate as shown; the sulfate forms part of *3-phosphoadenosine-5-phosphosulfate,* which serves as sulfate donor in conjugation reactions, for example, the formation of sulfates of catecholamines (page 614) and sulfated polysaccharides.

←

Figure 12.16
The metabolism of phenylalanine and tyrosine.
The details of the participation of tetrahydrobiopterin are shown.

CLIN. CORR. 12.5 THE HYPERPHENYLALANINEMIAS

The development of mass screening programs for blood phenylalanine levels, spurred by the discovery that a deficiency in the conversion of phenylalanine to tyrosine was the metabolic block in phenylketonuria (PKU), has led to the recognition of variants of PKU and also the realization that there are other causes of elevation of blood phenylalanine. Thus hyperphenylalaninemia and PKU are no longer automatically linked, although PKU still accounts for the vast majority of instances of hyperphenylalaninemia.

At present there have been described nine distinct types of hyperphenylalaninemia, many of them not associated with any disease state. Two of them (types 8 and 9) appear to be secondary to a primary tyrosinemia caused by a deficiency of *p*-hydroxyphenylpyruvate oxidase.

Type 1 is the classic PKU. This disorder was the first to be recognized as an inborn error of metabolism connected with mental retardation. The study of the epidemiology of PKU and its management has provided a model for the study of virtually all inborn errors of metabolism. Since mental retardation is a characteristic feature of PKU, this manifestation is generally looked for when dealing with inherited defects of amino acid metabolism. As indicated in another clinical correlation, the association is sometimes fortuitous.

PKU is characterized by the excretion of excessive amounts of phenylpyruvic acid in the urine. This compound is routinely screened for by the use of the ferric chloride test, which gives a green color with phenols and related compounds. However, this test is relatively non-

specific, since it gives a positive result with a number of compounds, including imidazolepyruvic acid (Clin. Corr. 12.3).

It is of interest to note that loss of pigmentation is a feature of PKU. The affected children have fair hair and blue eyes, and the normally dark areas of the brain (e.g., substantia nigra) are light. It has been shown that phenylalanine is a competitive inhibitor of tyrosinase, the enzyme which catalyzes the conversion of tyrosine to L-dopa in the formation of melanin. This, along with the actual reduction in the available tyrosine, can account for the loss of pigmentation.

The defective enzyme in PKU is phenylalanine hydroxylase, which catalyzes the conversion of phenylalanine to tyrosine. When this conversion is blocked or deficient, the normally minor pathway via transamination of phenylalanine to phenylpyruvic acid assumes major importance in the catabolism of phenylalanine. This minor pathway soon becomes overloaded, causing an elevation in blood phenylalanine. In addition to phenylpyruvic acid, other metabolites characteristic of this minor pathway, such as phenyllactic acid and phenylacetic acid also appear in the urine. A recently discovered hyperphenylalaninemia (type 5) is associated with a demonstrated defect in dihydropteridine reductase, the enzyme that catalyzes the regeneration of the tetrahydrobiopterin cofactor for phenylalanine hydroxylase. These patients were recognized as distinct because, unlike victims of PKU, their symptoms appeared shortly after birth and were unresponsive to the standard dietary treatment. In one patient, intravenous administration of tetrahydrobiopterin reduced the level of serum phenylalanine.

Tryptophan is broken down to *3-hydroxyanthranilic acid* (a precursor of niacin) and alanine in four steps (Figure 12.19). The first step is a cleavage of the pyrrole ring in the presence of oxygen to yield *N-formylkynurenine;* the reaction is catalyzed by *tryptophan-2,3-dioxygenase* and both atoms of oxygen are incorporated into the product. (Because of the opening of the pyrrole ring, the enzyme was formerly known as *tryptophan pyrrolase.*) The formyl group is removed, yielding *kynurenine,* which in turn is hydroxylated to yield *3-hydroxykynurenine,* which is the immediate precursor of alanine

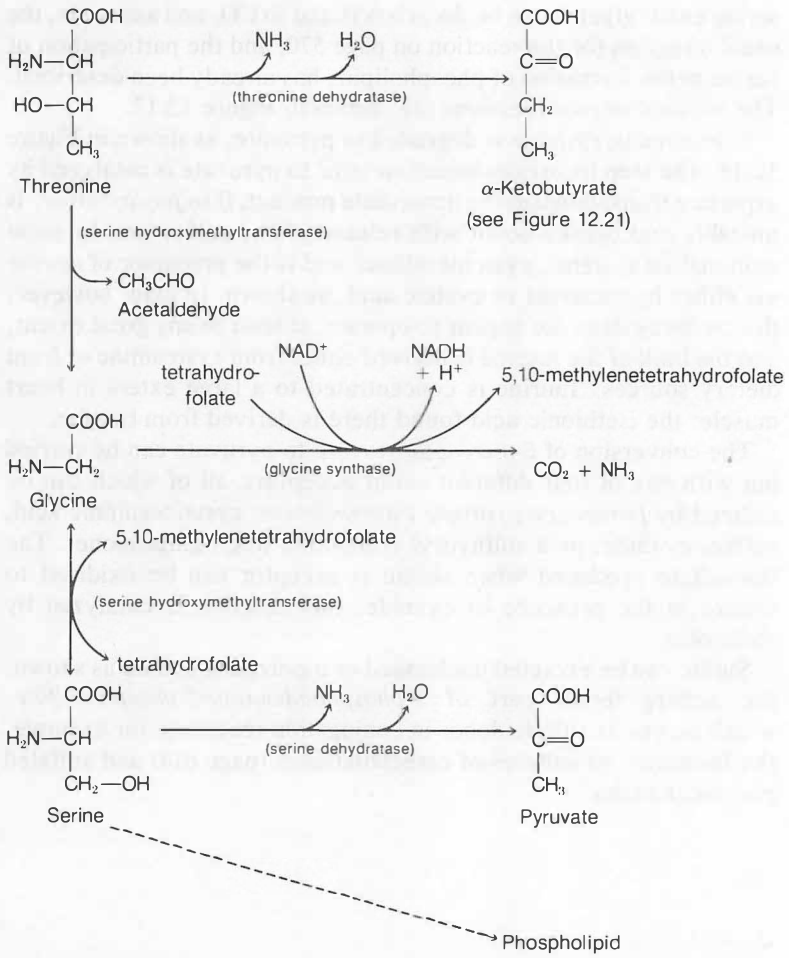


Figure 12.17
Catabolism of threonine, serine, and glycine.

CLIN. CORR. 12.6 DISORDERS OF TYROSINE OXIDATION

TYROSINOSIS AND THE TYROSINEMIAS

Tyrosinosis and the tyrosinemias have certain biochemical features in common, notably the excretion of large amounts of the early metabolites of tyrosine (*p*-hydroxyphenylpyruvate, *p*-hydroxyphenyllactate, and *p*-hydroxyphenylacetate) in the urine.

The information concerning tyrosinosis comes from a single patient whose symptoms were described in 1932. When the patient was on a low tyrosine diet or was fasting, *p*-hydroxyphenylpyruvate was the only metabolite excreted. At higher levels of tyrosine, *p*-hydroxyphenylpyruvate increased, and tyrosine was observed in the urine; at still higher levels, both *p*-hydroxyphenyllactate and 3,4-dihydroxyphenylalanine (dopa) were excreted.

The initial proposal for the metabolic block was a defect in *p*-hydroxyphenylpyruvate oxidase. However, if this were so, then *p*-hydroxyphenyllactate should have been the primary metabolite excreted. Also, this proposal does not account for the excretion of dopa at high levels of tyrosine ingestion. No enzyme studies were done on this patient; and since this is the only reported case, no new answers have been forthcoming.

The tyrosinemias all show elevated blood and urinary levels of tyrosine, in addition to the excretion of the other early metabolites of tyrosine. Transient tyrosinemia of the newborn appears to be due to delayed development of tyrosine transaminase and/or *p*-hydroxyphenylpyruvate oxidase. In some cases the administration of vitamin C has corrected the condition, presumably because of the requirement for this vitamin by *p*-hydroxyphenylpyruvate oxidase. Persistent tyrosinemia without liver or kidney disease has been reported in six patients, all of whom are retarded. Enzyme studies were done on only one patient, which turned up a deficiency in liver cytosol tyrosine transaminase; mitochrondrial transamination of tyrosine was normal. There also exists a hepatorenal type of tyrosinemia, characterized by cirrhosis, rickets, and a generalized aminoaciduria secondary to the renal disease. In this disorder, the excretion of *p*-hydroxyphenyllactate was quite high, indicating a deficiency in *p*-hydroxyphenylpyruvate oxidase; this was confirmed by direct enzyme assay.

ALCAPTONURIA

In contrast to many of the defects in amino acid metabolism, alcaptonuria has been known for several hundred years. It was first given its name in 1859, and in 1891 the compound responsible for the darkening of the urine, homogentisic acid (2,5-dihydroxyphenylacetate), was identified and its structure was established. Shortly thereafter, it was shown that tyrosine loading or a high-protein diet greatly increased the amount of homogentisic acid excreted by alcaptonuric patients. In 1908 Sir Archibald Garrod proposed the notion of inherited defects of metabolism, based partly on his studies of alcaptonuria. He suggested that an enzyme was defective or missing, and subsequent investigation, including direct enzyme assay, has borne him out.

The deficient enzyme in alcaptonuria is homogentisate oxidase, which catalyzes the splitting of the aromatic ring to yield maleylacetoacetate. In alcaptonuria, homogentisate is oxidized via a normally

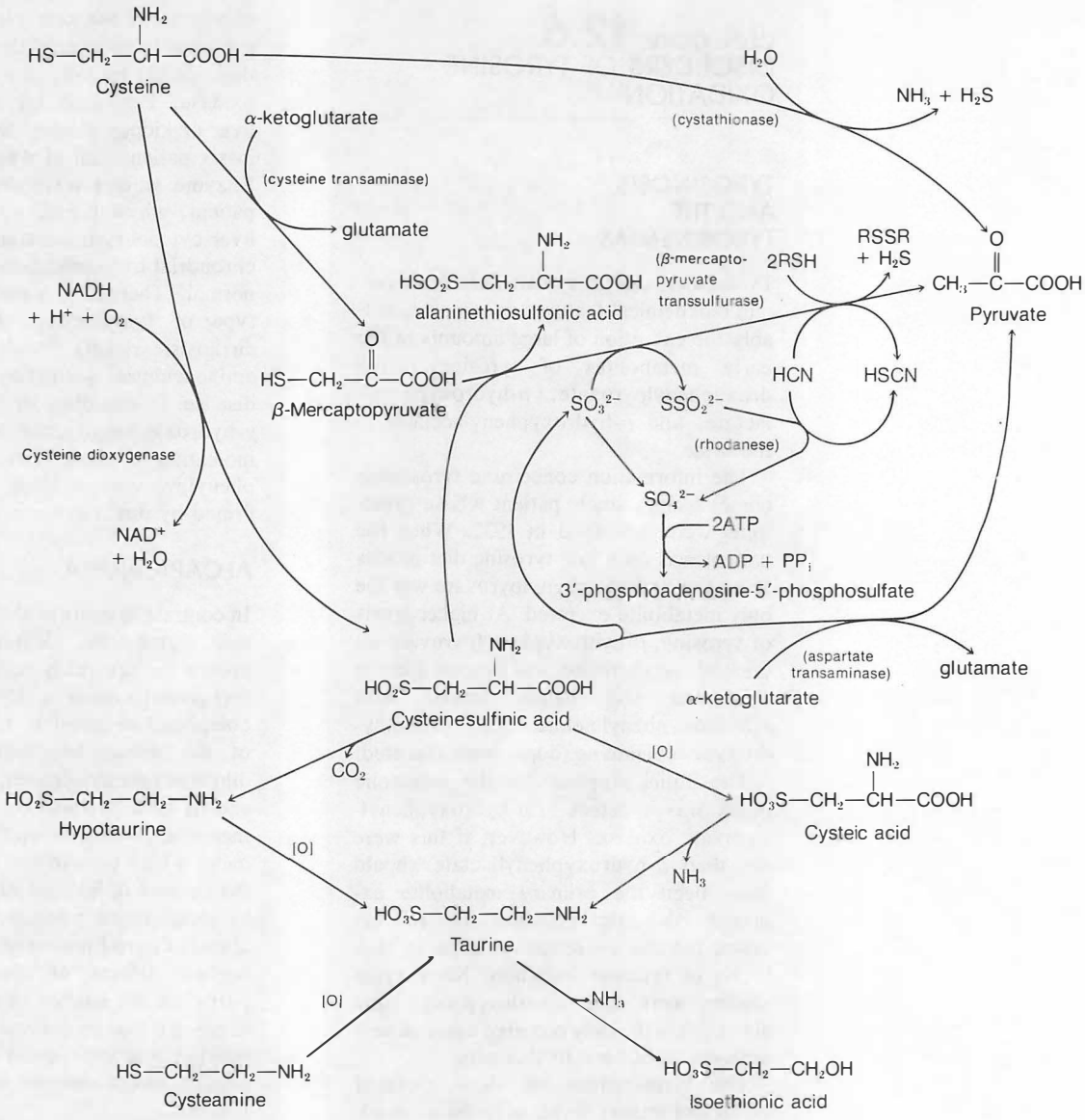


Figure 12.18
Catabolism of cysteine and formation of taurine.
As indicated, pyruvate can be formed directly from cysteine by the action of cystathionase; however, this appears to be a minor pathway.

and 3-hydroxyanthranilic acid; this last hydrolytic cleavage is catalyzed by *kynureninase*. The further metabolism of 3-hydroxyanthranilic acid is discussed below. It should be noted that the α-amino group of tryptophan appears as the α-amino group of alanine, and that the pyrrole nitrogen appears in the 3-hydroxyanthranilic acid. The further metabolism of tryptophan will be deferred to a later section.

Hydroxyproline undergoes oxidation and ring opening to yield γ-erythro-*hydroxyglutamate*, which is transaminated to yield *γ-hydroxy-α-ketoglutarate*. This compound is cleaved to give pyru-

minor pathway, yielding to the quinone, benzoquinone acetate, polymers of which are apparently responsible for the dark color. The darkening occurs only under alkaline conditions.

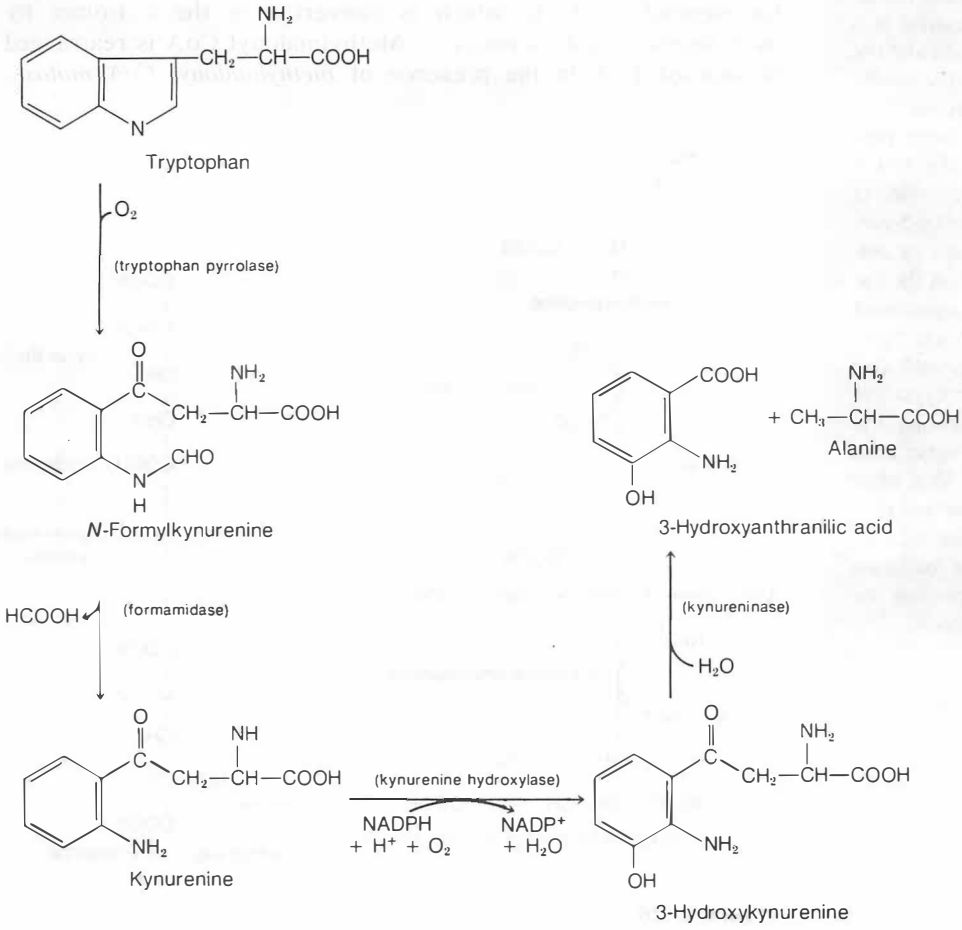


Figure 12.19
Formation of alanine and 3-hydroxyanthranilic acid from tryptophan.

CLIN. CORR. 12.7 HYPERHYDROXYPROLINEMIA

Hyperhydroxyprolinemia disorder is characterized by excessive amounts (15 to 50 times normal) of hydroxyproline in plasma. In the four cases reported, the patients exhibited a variety of symptoms, including mental retardation. Further investigation has failed to confirm a causal correction between mental retardation and the disorder. At present, therefore, the condition as such is regarded as harmless.

No enzymatic studies have been performed, but it is presumed that the defect is in the steps in which hydroxyproline is oxidized to Δ^1-pyrroline-3-hydroxy-5-carboxylic acid, rather than a defect in collagen metabolism. This is based on the absence of signs that would be associated with collagen metabolism, for example, disturbances of wound healing, and also on the observation that hydroxyproline loading resulted in levels of plasma hydroxyproline that were initially higher than normal and remained elevated long after the loading test. The observation that proline metabolism is normal in these individuals supports the notion that the oxidation of proline and that of hydroxyproline are catalyzed by two separate enzymes.

vate and glyoxylate; the latter is oxidized to oxalate and finally to CO_2 (Figure 12.20). (See Clin. Corr. 12.7.)

The Succinyl CoA Group: Methionine, Isoleucine, and Valine

As noted above, methionine is a precursor of cysteine via cystathionine; the other product of the cleavage is α-ketobutyric acid, which can also be generated from threonine. The α-ketobutyric acid is oxidatively decarboxylated in a reaction analogous to the decarboxylation of pyruvate or α-ketoglutarate. The reaction requires coenzyme A, and the product is *propionyl CoA*. This compound is carboxylated in a reaction requiring ATP; the product is D-*methylmalonyl CoA*, which is converted to the L isomer by *methylmalonyl CoA racemase*. L-Methylmalonyl CoA is rearranged to *succinyl CoA* in the presence of *methylmalonyl CoA mutase*.

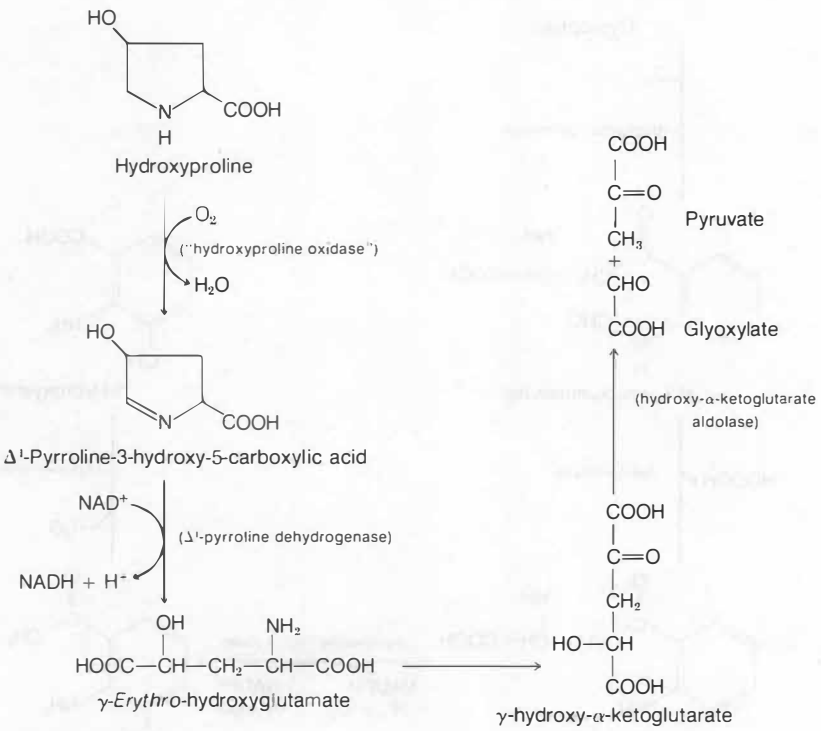


Figure 12.20
Catabolism of hydroxyproline.
Although it has not been fully characterized, hydroxyproline oxidase appears to be distinct from proline oxidase.

which requires the adenosyl form of vitamin B_{12} (cf. Figure 12.7). These reactions are shown in Figure 12.21. (See Clin. Corr. 12.8.)

The metabolism of *isoleucine* and *valine* follow rather similar steps. The first step for both is a transamination to the corresponding keto acid: the reactions are catalyzed by the *branched-chain amino acid transaminase,* with the concomitant formation of glutamate from α-ketoglutarate. The keto acids are decarboxylated in reactions

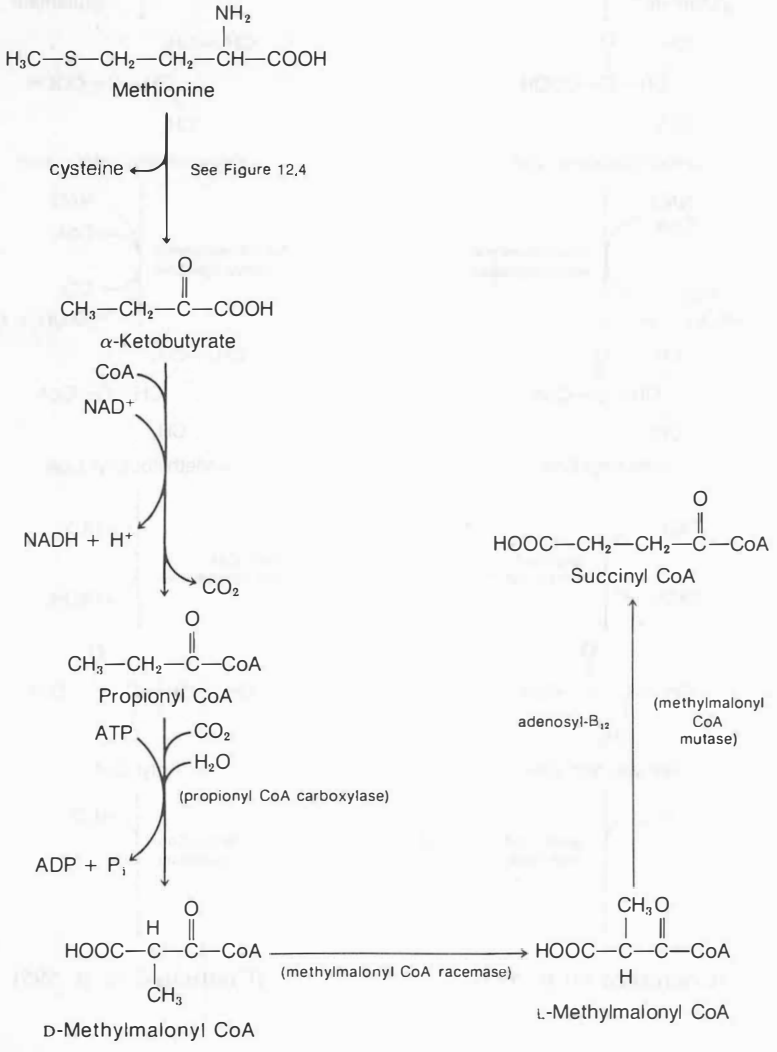


Figure 12.21
Formation of succinyl CoA from methionine.

CLIN. CORR. 12.8 DISORDERS OF PROPIONATE AND METHYLMALONATE METABOLISM

In view of the fact that the conversion of propionyl CoA to succinyl-CoA via methylmalonyl CoA is the last step in the metabolism of a number of compounds (Figure 12.23), it is not surprising that there are a variety of clinical manifestations that can be traced to defects in one or another step in this pathway. Since the mutase requires adenosylcobalamin, the discussion will include a consideration of some aspects of cobalamin metabolism. The cases will not be described in detail; instead, some common features will be noted and the discussion will focus on the nature of the metabolic defect.

All the patients described so far are children, many of them infants. Most of the patients are seriously ill; some display developmental defects of one sort or another. Sometimes the outcome is fatal, owing mostly to the profound ketoacidosis that is seen in most cases. Often the elevation of propionate and methylmalonate is accompanied by increases in blood levels of amino acids, mainly glycine; the mechanism responsible for this increase is not understood at present. The simultaneous occurrence of hyperglycinemia and ketosis has been called "ketotic hyperglycinemia." The hyperglycinemia is secondary to the defect in propionate and/or meth-

ylmalonate metabolism; it is thus to be distinguished from "nonketotic hyperglycinemia" which reflects a primary defect in glycine metabolism (cf. Clin. Corr. 12.2). However, one group of patients did not have ketoacidosis, but did have homocystinuria, cystathionuria, and hypomethioninemia.

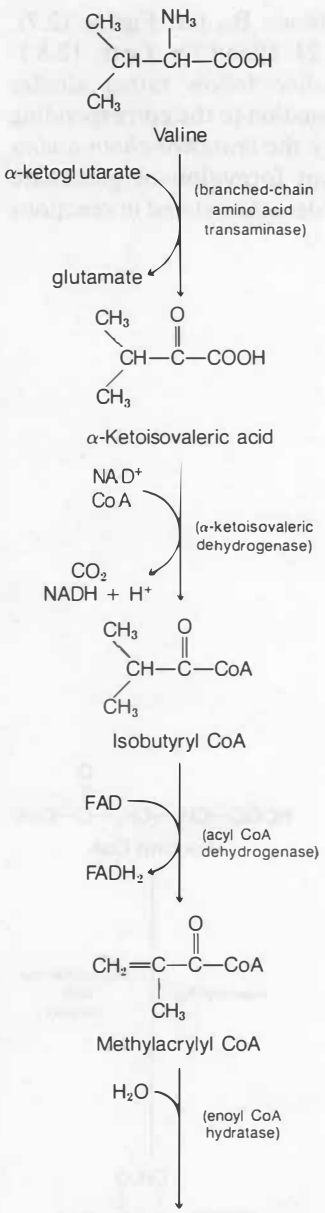
β-KETOTHIOLASE (ACETYL CoA ACYLTRANSFERASE) DEFICIENCY

Six cases of β-ketothiolase deficiency have been reported so far, four of which have been studied chemically. All the patients suffered episodes of ketoacidosis of varying intensity and frequency, ranging from repeated severe attacks to a single episode. The outcome was also highly variable, ranging from death to no apparent permanent effect.

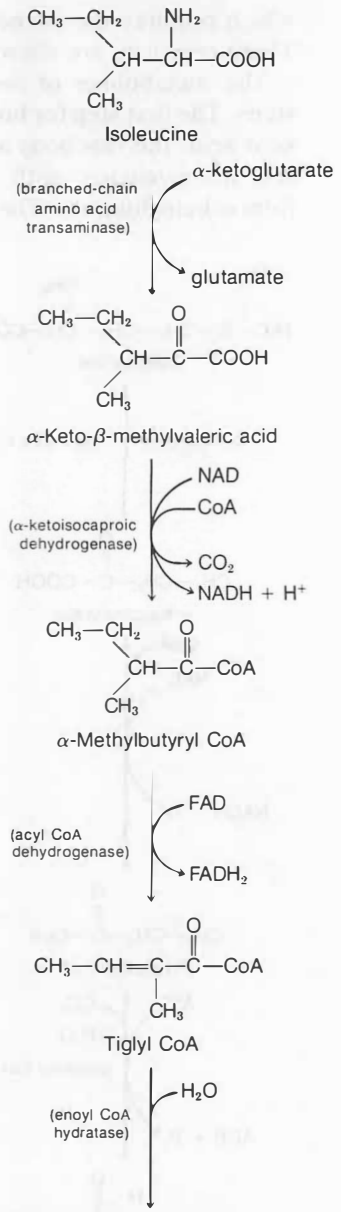
The urine of those patients studied chemically contained large quantities of α-methyl-β-hydroxybutyrate, α-methylacetoacetate, and butanone; the first two (actually, their CoA derivatives) are intermediates in the degradation of isoleucine to propionyl A. The activity of the enzyme was not measured directly, but since α-methylacetoacetyl CoA, together with CoA, are the substrates for the thiolase, a deficiency in the enzyme is presumed to be the defect.

PROPIONYL CoA CARBOXYLASE DEFICIENCY

Nearly 20 patients have been described with presumed or demonstrated inability to oxidize propionate or whose cells have been shown to have very low propionyl CoA carboxylase activity. Their symptoms include ketoacidosis, protein intolerance, and (often) developmental problems. Elevated serum propionate levels are a consistent feature. In one case, an infant who died on the fifth day of life, the con-



(Continued on p. 595)



(Continued on p. 595)

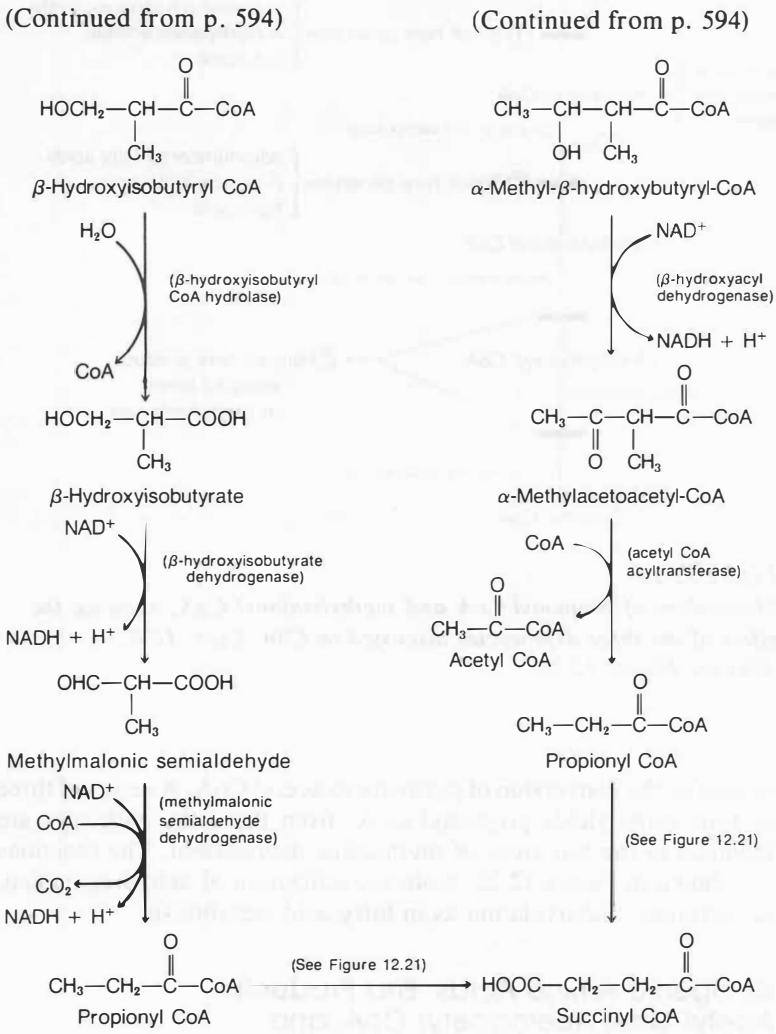


Figure 12.22
The metabolism of valine and isoleucine.
The two branched-chain keto acid dehydrogenases are probably parts of a system similar to pyruvate dehydrogenase, but only the first enzyme in each case is named here.

centration was more than 1,000 times normal.

In addition to the elevated propionate, other compounds were found in the serum. Among these are tiglic acid, an intermediate in isoleucine catabolism, and β-hydroxypropionate, a compound which appears in a minor pathway of propionate catabolism. Also observed was an accumulation of odd-numbered long-chain fatty acids in the liver. All these observations are consistent with a block in the conversion of propionyl CoA to methylmalonyl CoA. Tiglic acid and β-hydroxypropionate appear presumably because of the accumulation of propionyl CoA; the odd-numbered fatty acids arise because propionyl CoA is a "primer" for such fatty acid synthesis.

THE METHYLMALONIC ACIDEMIAS

Seventeen patients with methylmalonic acidemias have been described, all with the common feature of large amounts of methylmalonate in blood and urine; they fall into three more or less distinct groups. The first two groups had severe ketoacidosis; they were distinguished primarily by the fact that the first group was not responsive to cobalamine therapy, whereas the second group responded to pharmacological doses of either cyanocobalamin or adenosylcobalamin. The third group had no ketoacidosis but had homocystinuria, cystathionuria, and hypomethioninemia.

These findings were rationalized as follows: the first group had a primary defect in either methylmalonyl CoA racemase or methylmalonyl CoA mutase; the second group had a defect in the synthesis of adenosylcobalamine, thus reducing the activity of the mutase; and the third group had a primary defect in the retention and/or accumulation of cobalamin, thus reducing the availability of the vitamin for

synthesis of the coenzymes. Subsequently, one patient in the first group was shown to have the racemase deficiency, and a defect in adenosylcobalamin synthesis was demonstrated in the second group. No detailed studies were performed on the third group, so the explanation remains presumptive. However, the finding of altered levels of sulfur-containing amino acids in the third group is consistent with the proposed explanation of a defect in cobalamin accumulation, since this would also interfere with the remethylation of homocysteine to methionine (Figure 12.8).

The proposed and demonstrated enzyme deficiencies are sufficient to explain many of the findings, particularly the accumulation of certain metabolites; but several aspects of these disorders remain obscure. Hypoglycemia and hyperglycinemia are often observed as is hyperammonemia. In more cases the ketoacidosis is difficult to explain, since the levels of organic acids, although much higher than normal, do not seem sufficient to produce the observed drop in pH. Also, many of the hematological and neurological disturbances cannot be accounted for.

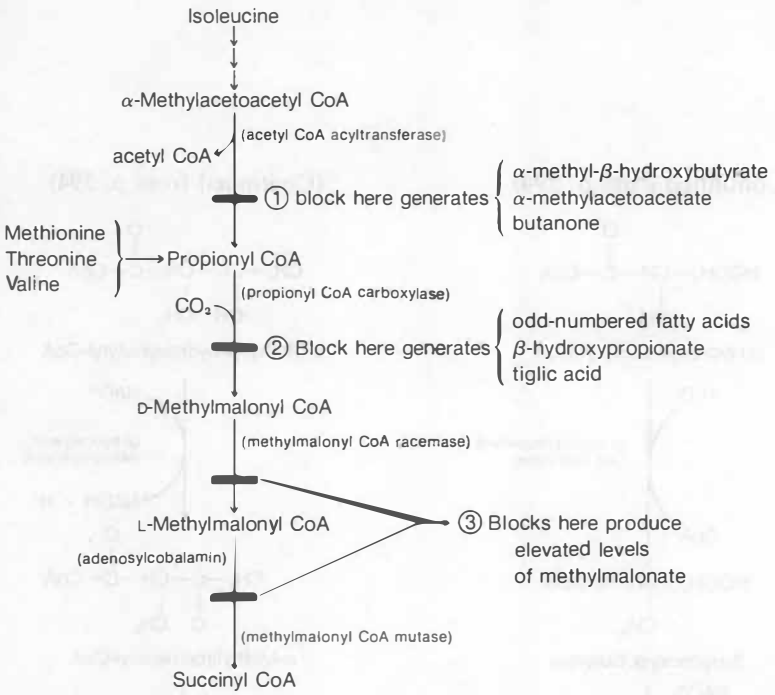


Figure 12.23
Metabolism of propionyl CoA and methylmalonyl CoA, showing the effect of the three deficiencies discussed in Clin. Corr. 12.8.
Also see Figure 12.21.

similar to the conversion of pyruvate to acetyl CoA. A series of three or four steps yields propionyl CoA; from there the pathways are identical to the last steps of methionine degradation. The reactions are shown in Figure 12.22. Note the utilization of dehydrogenation, rehydration, and oxidation as in fatty acid metabolism.

Ketogenic Amino Acids: End Products Acetyl CoA, Acetoacetyl CoA, and Acetoacetate

The Acetyl CoA Group: Isoleucine, Leucine, and Threonine

The metabolism of isoleucine has already been described in some detail (Figure 12.22). Acetyl CoA is produced in the step in which

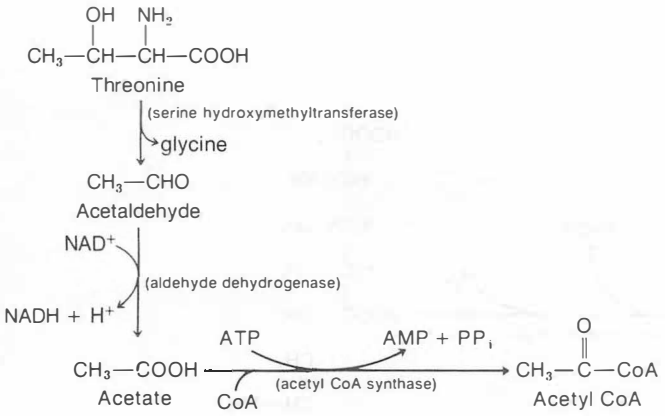


Figure 12.24
Formation of acetyl CoA from threonine.

α-methylacetoacetyl CoA is cleaved in the presence of another molecule of coenzyme A; the products are acetyl CoA and propionyl CoA (Figure 12.22).

As noted above, threonine is broken down to glycine and acetaldehyde by the action of serine hydroxymethyltransferase (Figure 12.17). The acetaldehyde is converted to acetyl CoA via acetate in the presence of coenzyme A; the first reaction is catalyzed by *acetaldehyde dehydrogenase* (Figure 12.24). The production of acetyl CoA from leucine will be described below.

The Acetoacetate/Acetoacetyl CoA Group: Phenylalanine, Tyrosine, Lysine, Leucine, and Tryptophan

The metabolism of phenylalanine and tyrosine has already been discussed. It remains only to note that the acetoacetate generated from the cleavage of fumarylacetoacetate is converted to acetoacetyl CoA. Succinyl CoA is the coenzyme A donor; this reaction is carried out only in extrahepatic tissue (See Figure 12.16).

The major route of *lysine* degradation begins with the condensation of lysine with α-ketoglutarate to produce *saccharopine;* this reaction is catalyzed by the first *saccharopine dehydrogenase* (lysine-forming). Saccharopine is then cleaved to glutamate and α-aminoadipic acid semialdehyde; the second *saccharopine dehydrogenase* (glutamate-forming) catalyzes this cleavage. These reactions and the subsequent steps from *α-aminoadipic acid semialdehyde* to acetoacetyl CoA are shown in Figure 12.25. They include steps similar to those found in fatty acid oxidation. Also shown is another pathway, involving an initial deamination to *α-keto-ε-aminocaproic acid,* cyclization to L-*pipecolic acid,* and subsequent ring opening to

CLIN. CORR. 12.9 THE HYPERLYSINEMIAS

Two types of hyperlysinemia have been reported: periodic hyperlysinemia with hyperammonemia, and persistent hyperlysinemia without hyperammonemia but with hyperlysinuria. The first type is represented by only one well-documented case, a female infant who developed symptoms of hyperammonemia. Despite the fact that the episodes of vomiting, spasticity, and coma were terminated successfully with appropriate therapy, she is noticeably retarded developmentally.

An enzyme study of biopsied liver tissue showed a deficiency in what the inves-

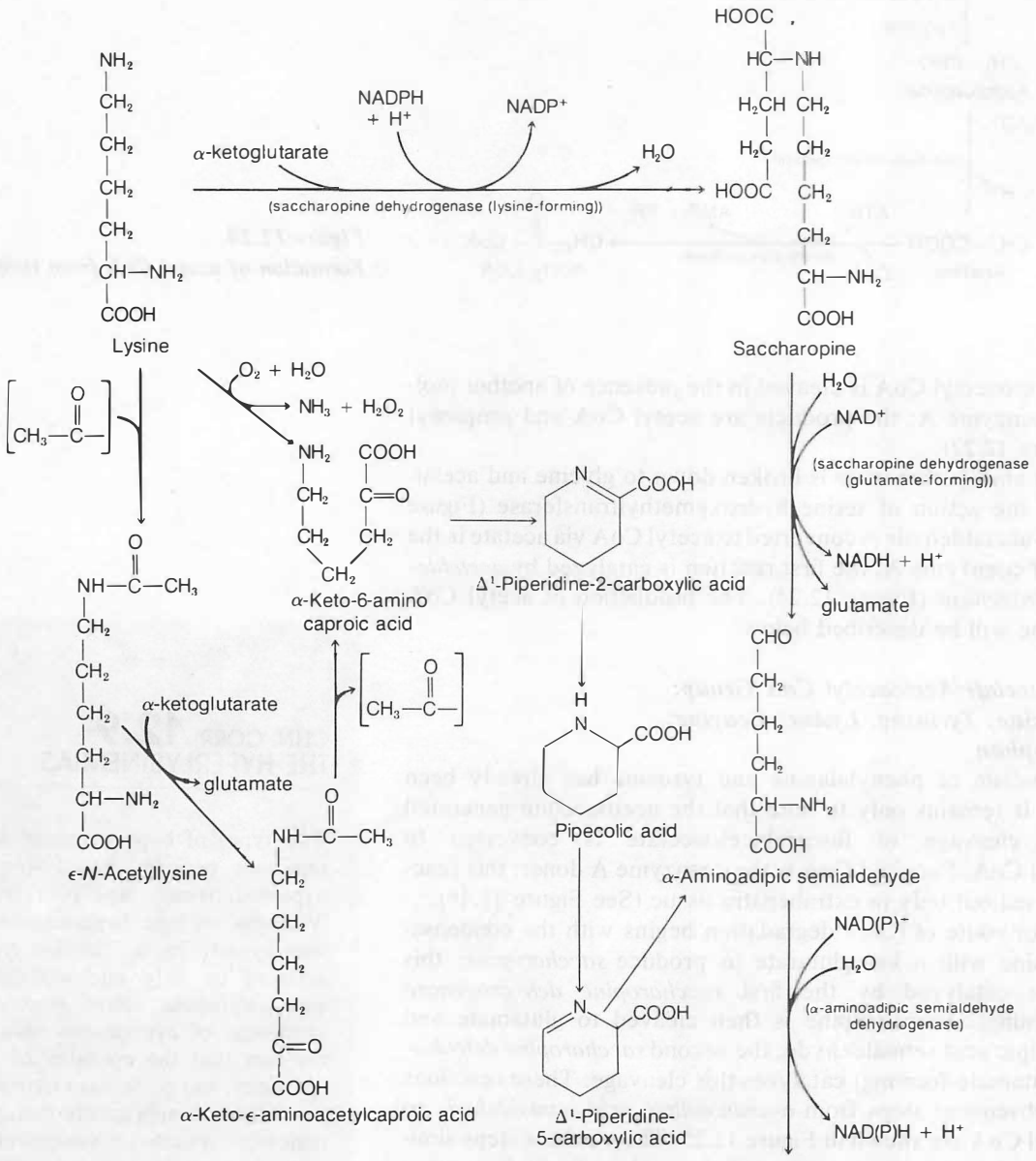
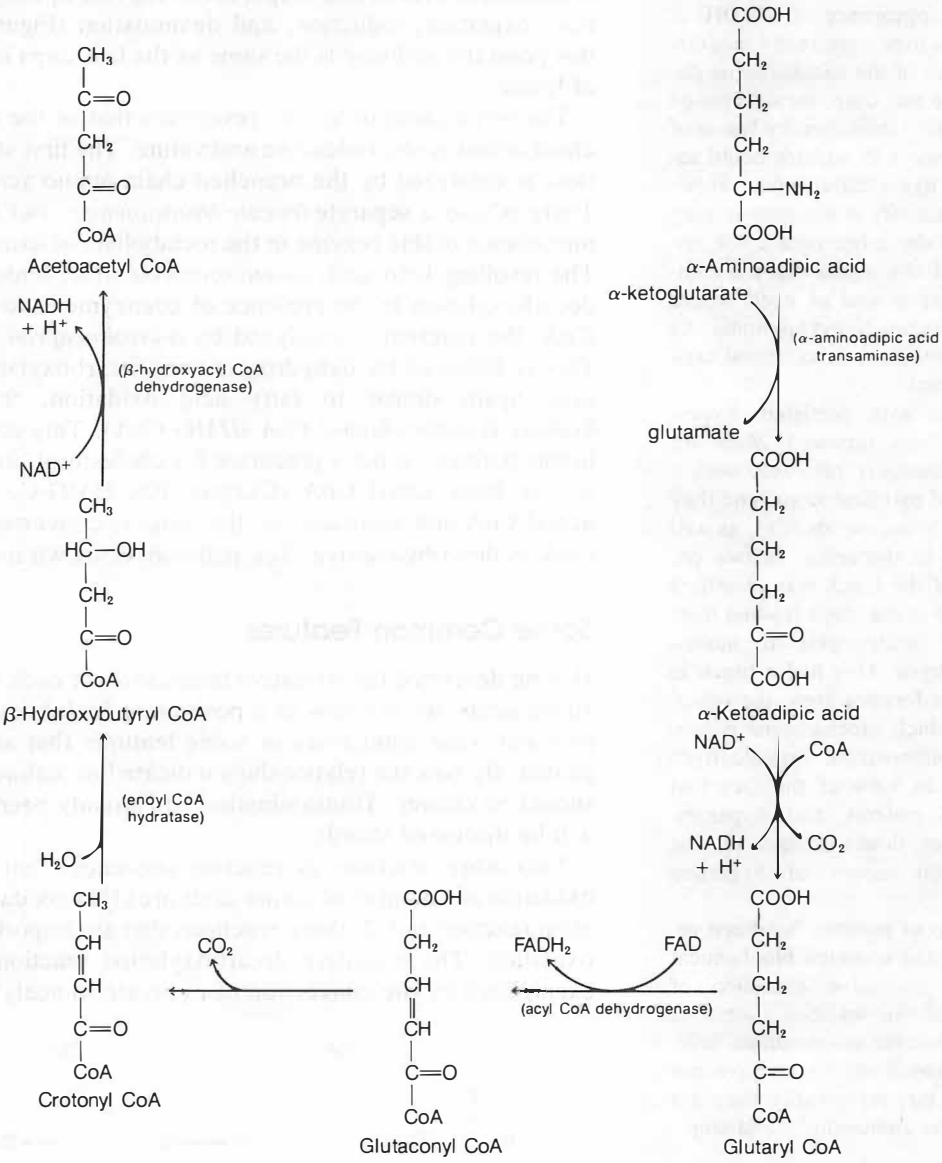


Figure 12.25
The metabolism of lysine.

(Continued on p. 599)

(Continued from p. 598)



tigators called a "lysine-deaminating" enzyme. Since they were using a crude supernatant as their enzyme preparation and following the course of the reaction by monitoring the appearance of NADH, it seems likely that they were really measuring either or both of the saccharopine dehydrogenases. In any case, they proposed that the competitive inhibition by lysine of arginase with respect to arginine could account for the hyperammonemia. However, since the activity of arginase is very high relative to the other urea cycle enzymes, it is hard to imagine that the competitive inhibition in and of itself would account for the excess blood ammonia. As with other hyperammonemias, blood urea levels were normal.

Seven patients with persistent hyperlysinemia have been reported. Most are physically and mentally retarded, with a wide spectrum of manifestations, and they excrete various lysine metabolites, as well as lysine itself, in the urine. In two patients the site of the block was identified and shown to be in the steps leading from lysine through saccharopine to aminoadipic semialdehyde. One had a block in the saccharopine-forming step; the other, in the step in which saccharopine is broken down to aminoadipic semialdehyde and glutamate. In view of the fact that none of these patients had hyperammonemia, further doubt is cast on the lysine-competition theory of hyperammonemia.

Another group of patients has been reported in which the common biochemical feature is the excessive excretion of aminoadipic acid. In addition, some of these patients excrete α-ketoadipic acid. The metabolic block or blocks have not been identified, but presumably they are subsequent to the aminoadipic acid step.

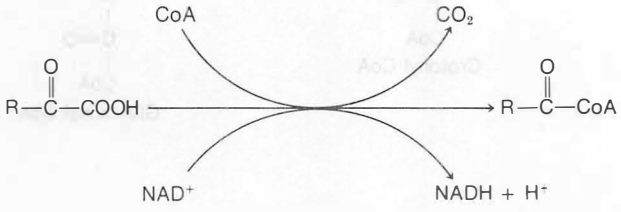
α-aminoadipic acid semialdehyde. The evidence for the existence of this pathway is not conclusive; in any event, it does not appear to be a primary pathway. As noted above, *tryptophan* is metabolized to alanine and 3-hydroxyanthranilic acid; the latter is converted to α-ketoadipic acid in four steps, involving ring opening, decarboxylation, oxidation, reduction, and deamination (Figure 12.26). From this point the pathway is the same as the last steps in the catabolism of lysine.

The metabolism of *leucine* resembles that of the other branched-chain amino acids, isoleucine and valine. The first step, transamination, is catalyzed by the branched-chain amino acid transaminase. There is also a separate *leucine transaminase*, but the quantitative importance of this enzyme in the metabolism of leucine is uncertain. The resulting keto acid, *α-ketoisocaproic acid*, undergoes oxidative decarboxylation in the presence of coenzyme A to yield isovaleryl CoA; the reaction is catalyzed by *α-ketoisocaproic dehydrogenase*. This is followed by dehydrogenation, recarboxylation, and hydration, again similar to fatty acid oxidation; the result is *β-hydroxy-β-methylglutaryl CoA* (*HMG CoA*). This compound arising in this pathway is not a precursor for cholesterol, in contrast to that arising from acetyl CoA (Chapter 10). HMG-CoA is cleaved to acetyl CoA and acetoacetate; the latter is converted to acetoacetyl CoA as described above. The pathway is shown in Figure 12.27.

Some Common Features

Having described the oxidative breakdown of each of the individual amino acids, we are now in a position to look back and attempt to pick out some similarities or some features that are shared or repeated. By now the relationships indicated in outline in Figure 12.13 should be clearer. Transamination has already been mentioned and will be discussed shortly.

Two other reactions or reaction sequences that are part of the oxidation of a number of amino acids are (1) an oxidative decarboxylation reaction and (2) three reactions that are important in fatty acid oxidation. The oxidative decarboxylation reaction is of the type exemplified by the conversion of pyruvate to acetyl CoA:



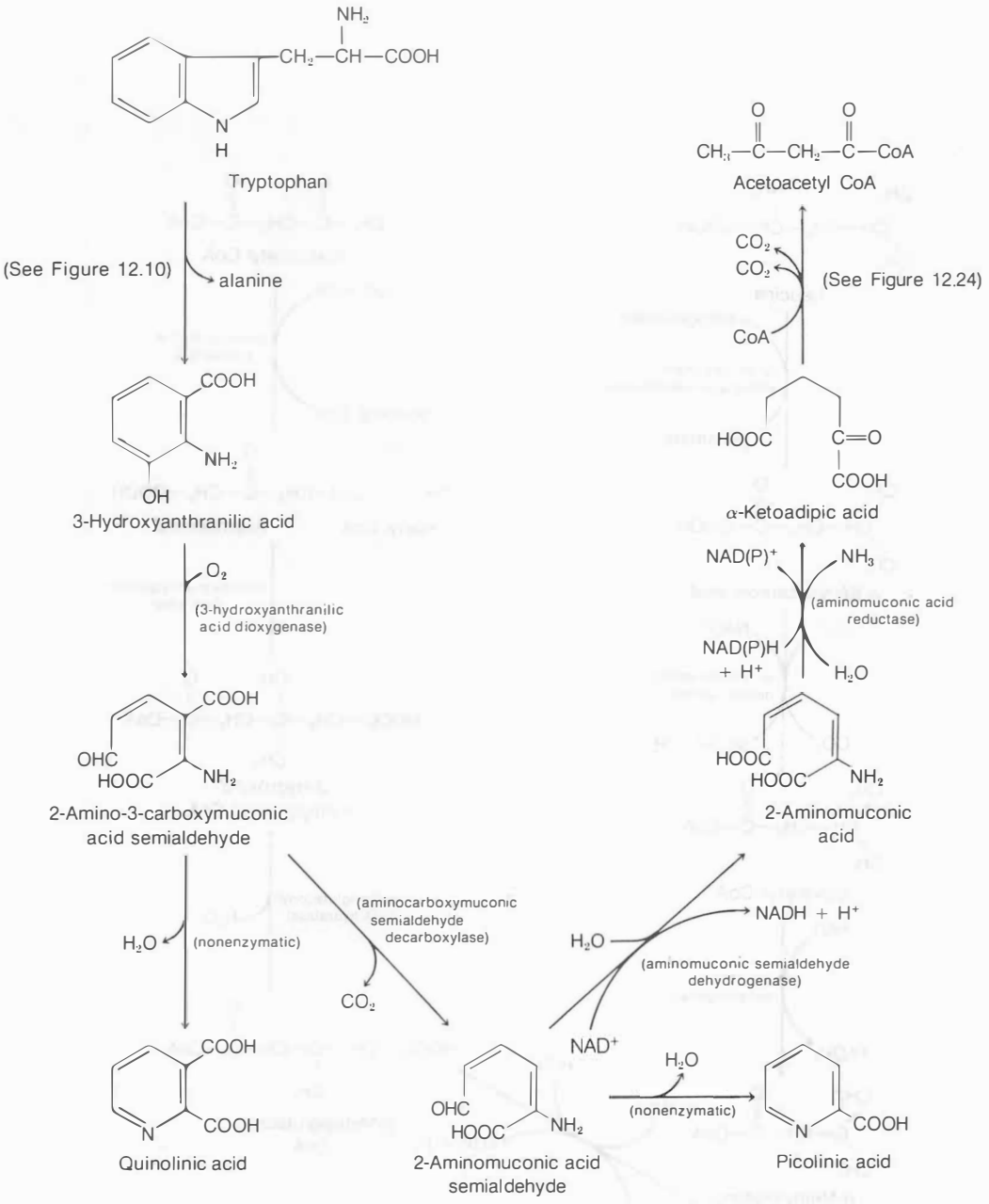


Figure 12.26
Formation of acetoacetyl CoA from tryptophan.

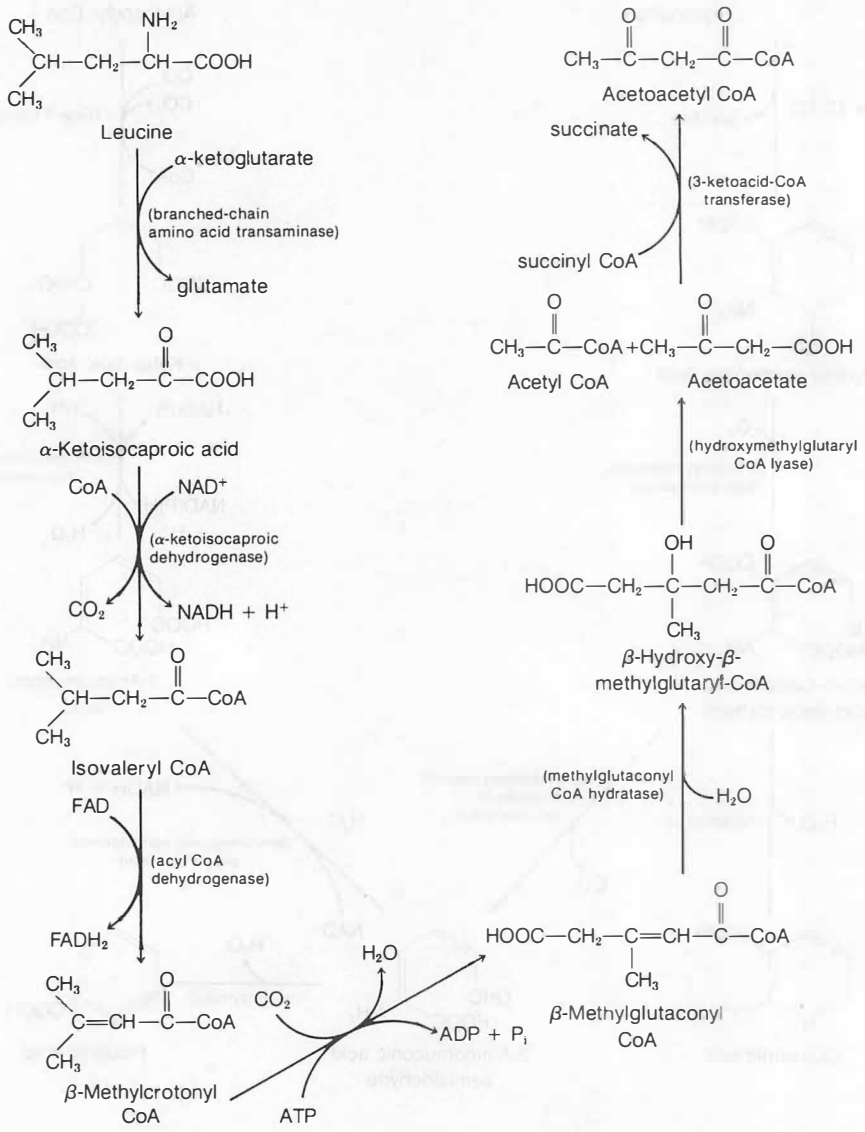
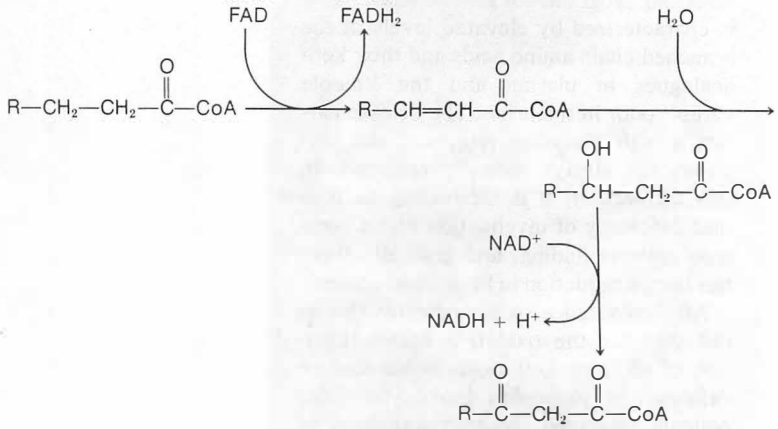


Figure 12.27
The metabolism of leucine.

The individual reactions of this type that occur in the oxidation of amino acids are shown in Table 12.3. Note again the methionine–threonine and lysine–tryptophan pairs.

The reaction sequence shown here



is the heart of fatty acid oxidation (Chapter 9). This sequence appears either unchanged or slightly modified in the breakdown of all the amino acids in Table 12.3, except for methionine and threonine. For isoleucine, the sequence is followed strictly, beginning with α-methylbutyryl CoA. The sequence is interrupted for valine after the hydration step by the removal of CoA from β-hydroxyisobutyryl CoA. For lysine and tryptophan, it is interrupted after the dehydrogenation step by removal of the δ-carboxyl group from glutaconyl CoA; it is interrupted after the same step for leucine, but for a

Table 12.3 Oxidative Decarboxylation Reactions in Amino Acid Metabolism

Parent Amino Acid	*Substrate*	*Product*
Methionine, Threonine	α-Ketobutyrate	Propionyl CoA
Valine	Methylmalonyl semialdehyde	Propionyl CoA
Valine	α-Ketoisovalerate[a]	Isobutyryl CoA
Isoleucine	α-Keto-β-methylvalerate[a]	α-Methylbutyryl CoA
Leucine	α-Ketoisocaproate[a]	Isovaleryl CoA
Lysine, Tryptophan	α-Ketoadipate[a]	Glutaryl CoA

[a] Generated from transamination. See Table 12.4.

CLIN. CORR. 12.10 ABNORMALITIES OF THE METABOLISM OF BRANCHED-CHAIN AMINO ACIDS

The group of disorders diagnosed as abnormalities of the metabolism of branched-chain amino acids involves the initial steps of the metabolism of the branched-chain amino acids—leucine, isoleucine, and valine. As a group, the disorders are quite serious and often fatal, and the survivors are usually severely retarded, both mentally and physically.

At present, there are four known defects in this group: hypervalinemia, maple syrup urine disease (branched-chain ketonuria), isovaleric acidemia, and β-methylcrotonic aciduria. The last two reflect defects in two different steps of leucine metabolism; maple syrup urine disease involves all three amino acids. Refer to Figures 12.22 and 12.27 for the metabolism of the branched-chain amino acids.

HYPERVALINEMIA

One case of hypervalinemia has been described. The diagnosis of a defect in the transamination of valine to α-ketoisovaleric acid was suspected because of the absence of ketoaciduria, and confirmed by assay of enzyme activity in leukocytes, which showed the expected deficiency in the conversion of valine to α-ketoisovaleric acid. Isoleucine and leucine were apparently handled normally. This case is of interest because the existence of a specific transaminase for valine runs counter to the accepted idea that a single enzyme catalyzes the transamination of all three amino acids.

MAPLE SYRUP URINE DISEASE

Maple syrup urine disease is now well enough recognized to be included in screening programs for genetic diseases. It is characterized by elevated levels of the branched-chain amino acids and their keto analogues in plasma and the "maple syrup" odor in the urine. The disease carries a high mortality rate, and the survivors are always mentally retarded. In this connection, it is interesting to note that deficiency of myelination was a common autopsy finding, and generally there has been a reduction in brain lipid content.

All the evidence so far indicates that in this disorder the oxidative decarboxylation of all three keto acids is blocked or deficient; in particular, leukocytes from patients catalyzed the transamination of the amino acids normally, but decarboxylation of the keto acids was poor or absent.

Attempts to identify a toxic agent which interferes with brain function point more or less to leucine: clinically, the acute symptoms seem to be more directly related to excess of this amino acid than to excess of the other two. In cultures of cerebellar cells, only α-ketoisocaproic acid (generated from leucine by transamination) interfered with myelination, and this keto acid also reduced oxygen consumption in brain slices, an effect also produced by valine. The metabolites responsible for the maple syrup odor are unknown.

ISOVALERIC ACIDEMIA

Isovaleric acidemia, like maple syrup urine diesase, also carries a high mortality rate; however, the survivors do not appear to be as severely retarded developmentally. A characteristic odor, described variously as "cheesy" or "like sweaty feet," often accompanies the disorder. This odor is

characteristic of branched-chain compounds.

The defect is presumed to be in the step in which isovaleryl CoA is converted to β-methylcrotonyl CoA. In addition to isovaleric acid in the plasma of these patients, β-hydroxyisovaleric acid was found in the urine. Leukocytes and skin fibroblasts show a severe deficiency in the oxidation of isovaleric acid to CO_2, but a specific defect in isovaleryl CoA dehydrogenase has not yet been found.

β-METHYLCROTONIC ACIDURIA

The patients in this group excreted β-methylcrotonic acid and β-hydroxyisovaleric acid in large quantities. No enzyme studies were performed, but the defect was presumed to be in β-methylcrotonyl CoA carboxylase. This was supported by the finding that administration of biotin caused a dramatic improvement. The carboxylase is biotin-dependent, and it may be that the defect involves biotin, rather than the enzyme protein itself.

CLIN. CORR. 12.11 HARTNUP DISEASE

Strictly speaking, Hartnup disease is not a disorder of amino acid metabolism as such; rather, it is a disturbance of the renal and intestinal reabsorption mechanisms for the neutral amino acids, resulting in excessive excretion of these amino acids in the urine. Clinically, the patients exhibited symptoms, particularly rash and neurological and mental disturbances, which were strongly suggestive of pellagra. Since Hartnup patients respond very well to oral nicotinamide therapy, it was concluded that the symptoms were due to a relative deficiency of tryptophan, one of the group of neutral amino acids whose reabsorption is interfered with. The pellagra-like symptoms can be rationalized on the basis of the defect in intestinal absorption of tryptophan, which is metabolized by intestinal bacteria to a variety of products, including indole. Under conditions of defective intestinal absorption, indole accumulates and interferes with the breakdown of tryptophan to *N*-formylkynurenine.

It will be recalled (see text) that 3-hydroxyanthranilic acid, a precursor of niacin, is derived from tryptophan. The amount of niacin produced by this route is quite small under these conditions, but it may be that there is some local specialized requirement for nicotinic acid, which cannot be met by preformed exogenous sources. The symptoms do not appear unless the individual subjected to situations in which dietary nicotinamide is in short supply. Growth is one of these situations, and this might account for the fact that affected individuals deteriorate clinically during childhood and then improve as adults.

Table 12.4 Removal of Amino Groups from Amino Acids

REMOVAL BY TRANSAMINATION		
Parent Amino Acid	*Compound Undergoing Transamination*	*Keto Product*
Cysteine	Cysteine	β-Mercaptopyruvate
Cysteine	Cysteinesulfinate	Pyruvate, with loss of SO_3
Lysine	α-Aminoadipate	α-Ketoadipate
Valine	Valine	α-Ketoisovalerate
Leucine	Leucine	α-Ketoisocaproate
Isoleucine	Isoleucine	α-Keto-β-methylvalerate
Phenylalanine-tyrosine	Tyrosine	*p*-Hydroxyphenylpyruvate
Hydroxyproline	γ-*erythro*-Hydroxyglutamate	γ-Hydroxy-α-ketoglutarate

REMOVAL BY REACTIONS OTHER THAN TRANSAMINATION		
Parent Amino Acid	*Compound Undergoing Reaction*	*Product(s)*
Methionine	Cystathionine	α-Ketobutyrate + NH_3 + cysteine
Glycine	Glycine	CO_2 + NH_3
Serine	Serine	Pyruvate
Threonine	Threonine	α-Ketobutyrate
Histidine	Histidine	Urocanic acid + NH_3

carboxylation, to generate β-methylglutaconyl CoA from β-methyl-crotonyl CoA. Also, for leucine, the oxidation step is actually a cleavage.

It should be noted that oxidative decarboxylation (as described above) and the fatty acid oxidation sequence appear only in the breakdown of the longer chain aliphatic amino acids (except arginine), plus threonine and tryptophan. The other short-chain aliphatic amino acids (glycine, alanine, aspartate, etc.) are more or less interchangeable; and threonine has other routes of degradation available besides conversion to α-ketobutyrate. The degradative pathways for the aromatic amino acids are essentially unique to each acid, with the exception of tryptophan.

As mentioned in Chapter 11, nearly every amino acid undergoes

transamination at some point as a means of removal of the amino group. Alanine, aspartate, and glutamate constitute a special group whose members are interconvertible by transamination. In addition, proline and arginine are first converted to glutamate before undergoing transamination. The α-amino group of tryptophan is not removed until the molecule is separated into 3-hydroxyanthranilic acid and alanine; the latter carries the original α-amino group.

In Table 12.4 are shown the reactions by which the rest of the amino acids lose their amino groups. The majority are converted by transamination into something other than α-ketoglutarate, pyruvate, or oxalacetate. Most of the others undergo direct deamination.

Gluconeogenesis and the Glycogenic Amino Acids: The Alanine–Glucose Cycle

As indicated above, the glycogenic amino acids, those that produce a net synthesis of glucose, yield one of five end products: pyruvate, α-ketoglutarate, succinyl CoA, fumarate, and oxalacetate. In gluconeogenesis, pyruvate is converted to phospho*enol*pyruvate via oxalacetate (Chapter 7). Obviously, then, oxalacetate itself, plus the other three TCA cycle intermediates, are capable of being converted to phospho*enol*pyruvate through the action of phospho*enol*-pyruvate carboxykinase. The relationships are shown in Figure 12.28.

Of special interest is alanine. Quantitatively, it is the most important amino acid taken up by the liver from the peripheral tissues, particularly from skeletal muscle; thus, it is a major participant in the interorgan transfer of nitrogen. Alanine constitutes no more than 7–10% of muscle protein, yet accounts for more than 30% of the α-amino nitrogen given to liver from that source. It has been suggested that a large part of this alanine is synthesized by transamination in the muscle, and that there is a cyclic process between liver and muscle involving alanine and glucose. This "alanine–glucose" cycle is shown in Figure 12.29.

Amino Acids as Precursors of Other Nitrogenous Compounds

Aspartate, Glycine, and Glutamine

These three amino acids play important roles in the biosynthesis of purines and pyrimidines. Glycine is incorporated whole into the purine ring, while aspartate and glutamine each contribute nitrogen. Also, aspartate ultimately supplies three carbons and a nitrogen to the pyrimidine ring (see Chapter 13 for details).

Besides its incorporation into the purine ring, glycine also partici-

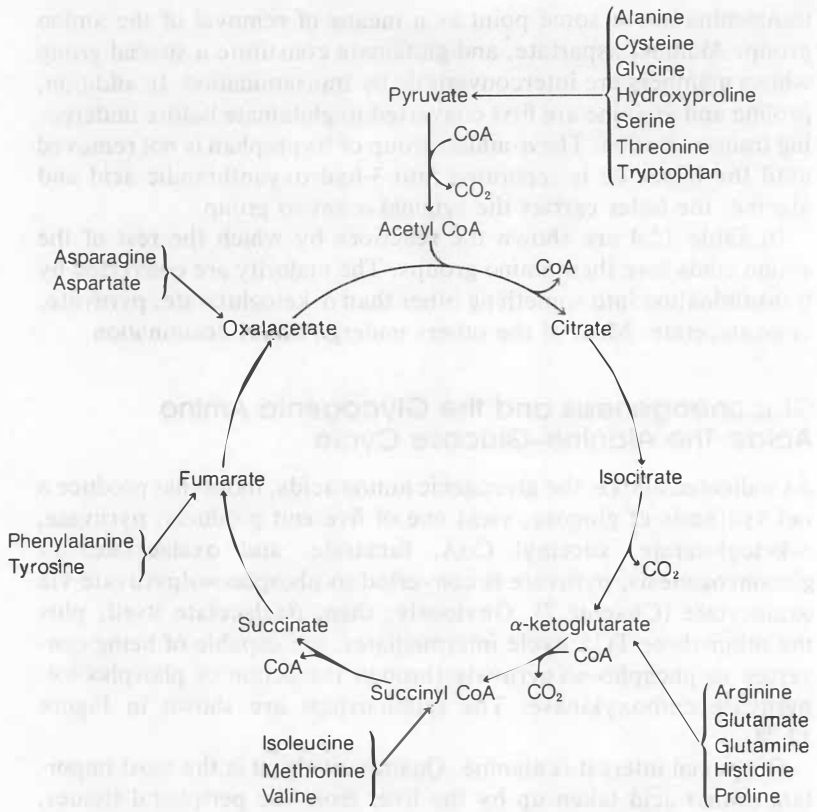
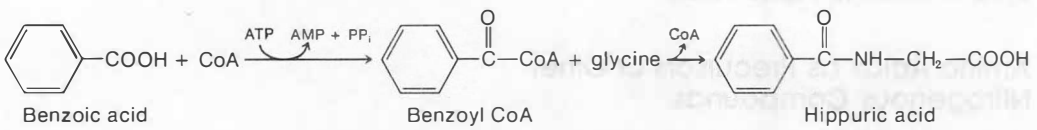


Figure 12.28
The relationship of the glycogenic amino acids to the TCA cycle.

pates in the synthesis of porphyrins. It condenses with succinyl CoA to form *δ-aminolevulinic acid*, the precursor of porphobilinogen (Chapter 22). Another function of glycine is that of detoxification by forming conjugates, as illustrated by the formation of *hippuric acid* (*N*-benzoylglycine) from benzoic acid:



Hippuric acid is excreted in the urine. The formation of glycine conjugates from CoA esters is catalyzed by *glycine acyltransferase.* Cholic acid and its derivatives are excreted similarly (glycocholic acids), or as derivatives of taurine (taurocholic acids).

In the formation of creatinine, glycine and arginine react to form ornithine and *guanidinoacetic acid;* the reaction is catalyzed by

glycine amidinotransferase. In human beings the highest activity of this enzyme is in the pancreas. In the next step, guanidinoacetic acid is *N*-methylated to yield *creatine;* the methyl donor is *S*-adenosylmethionine, and the reaction is catalyzed by *guanidinoacetate methyltransferase,* a liver enzyme. Excess production of creatine results in a fatty liver. In this connection, it has been suggested that ornithine can serve as an acceptor for the amidino group in place of glycine; this would have the effect of controlling the synthesis of creatine. The reactions are shown in Figure 12.30.

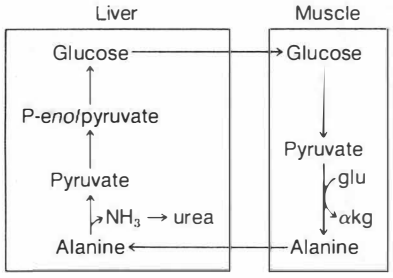
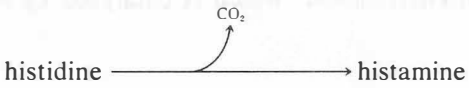


Figure 12.29
The "alanine-glucose" cycle.

Histidine

Histidine is the precursor of *histamine,* a powerful vasodilator secreted by mast cells in allergic reactions and in response to trauma. The production of histamine is catalyzed by a specific *histidine decarboxylase:*



The further metabolism of histamine involves deamination to *imidazolylacetaldehyde,* oxidation to *imidazolylacetic acid,* and formation and excretion of the ribonucleotide of imidazolylacetate. The sequence is shown in Figure 12.31.

The reaction in which histamine is formed is one of six direct decarboxylations of amino acids that we will consider in which a primary amine is generated. This type of decarboxylation is not to be confused with the CoA-dependent oxidative decarboxylation reac-

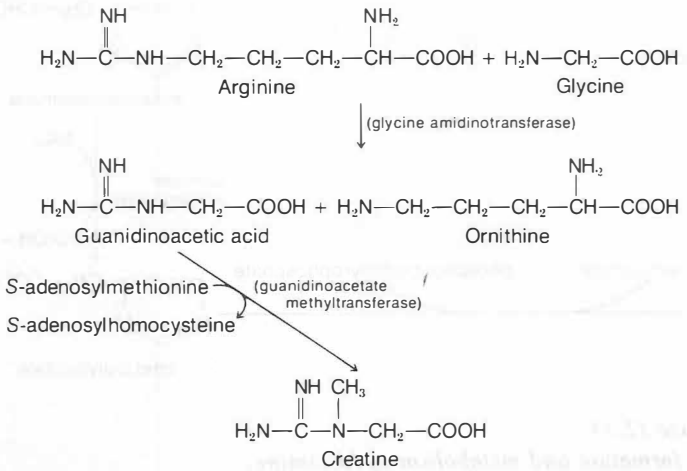


Figure 12.30
The synthesis of creatine.

tions described earlier which take place during the oxidative breakdown of amino acids. The other decarboxylation reactions will be introduced in the appropriate contexts. Table 12.5 summarizes these reactions.

Tryptophan

Tryptophan is a precursor of *nicotinic acid* via 3-hydroxyanthranilic acid and *2-amino-3-carboxymuconic acid semialdehyde* (Figure 12.26). The sequence begins with the spontaneous closure of the latter compound to produce *quinolinic acid,* which undergoes decarboxylation and condensation with ribose-5-phosphate. The resulting compound is *nicotinic acid ribonucleotide,* which is the direct precursor of NAD. The conversion of tryptophan to quinolinic acid and nicotinic acid ribonucleotide is shown in Figure 12.32.

Tryptophan is also the precursor of the neurotransmitter *serotonin* (5-hydroxytryptamine) (Figure 12.32). The first step is the formation of *5-hydroxytryptophan,* which is catalyzed by *tryptophan hy-*

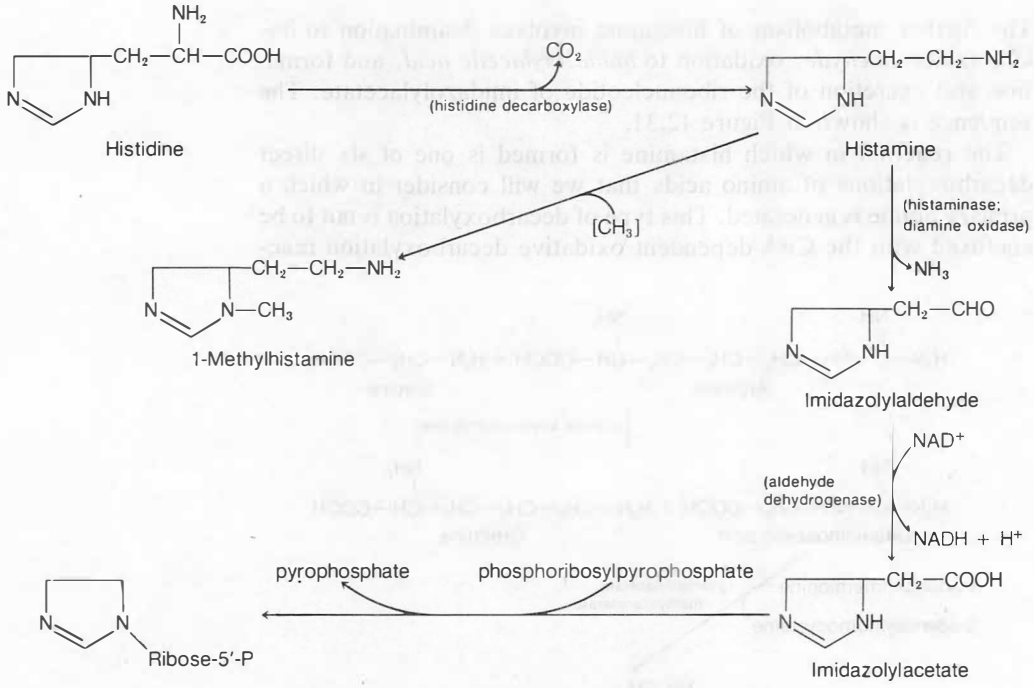


Figure 12.31
The formation and metabolism of histamine.

Table 12.5 Direct Decarboxylations of Amino Acids

Parent Amino Acid	*Substrate*	*Product*	*Biosynthetic End Product(s)*
Histidine	Histidine	Histamine	Histamine
Glutamate	Glutamate	γ-Aminobutyrate	γ-Aminobutyrate
Ornithine	Ornithine	Putrescine	Spermidine and spermine
S-Adenosyl-methionine	*S*-Adenosyl-methionine	*S*-Adenosyl-methylthio-propylamine	Spermidine and spermine
Tryptophan	5-Hydroxy-tryptophan	Serotonin (5-hydroxy-tryptamine)	Serotonin
Phenylalanine/ tyrosine	L-Dopa	Dopamine	Dopamine, norepinephrine, epinephrine

droxylase, an enzyme whose mode of action is quite similar to phenylalanine hydroxylase (see above). Tryptophan hydroxylase is found mainly in brain, especially in the pineal gland.

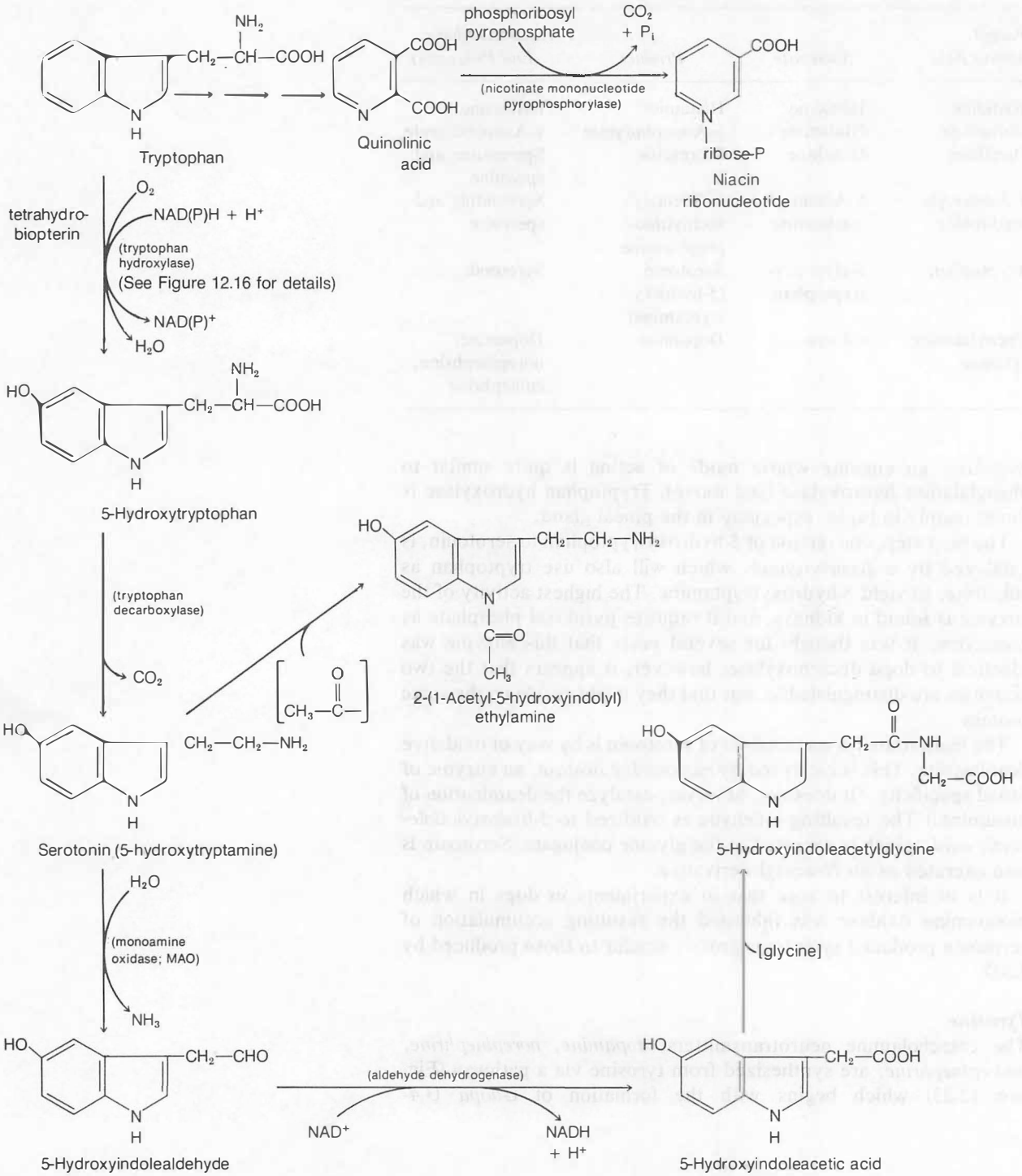
The next step, conversion of 5-hydroxytryptophan to serotonin, is catalyzed by a *decarboxylase*, which will also use tryptophan as substrate, to yield 5-hydroxytryptamine. The highest activity of the enzyme is found in kidneys, and it requires pyridoxal phosphate as coenzyme. It was thought for several years that this enzyme was identical to dopa decarboxylase; however, it appears that the two activities are distinguishable, but that they might reside on the same protein.

The main route for metabolism of serotonin is by way of oxidative deamination. This is catalyzed by *monoamine oxidase*, an enzyme of broad specificity. (It does not, however, catalyze the deamination of histamine.) The resulting aldehyde is oxidized to *5-hydroxyindole-acetic acid*, which is excreted as the glycine conjugate. Serotonin is also excreted as an *N*-acetyl derivative.

It is of interest to note that in experiments in dogs in which monoamine oxidase was inhibited the resulting accumulation of serotonin produced symptoms grossly similar to those produced by LSD.

Tyrosine

The catecholamine neurotransmitters, *dopamine*, *norepinephrine*, and *epinephrine*, are synthesized from tyrosine via a pathway (Figure 12.33) which begins with the formation of L-*dopa* (3,4-


Tryptophan
Quinolinic acid
phosphoribosyl pyrophosphate
CO_2 + P_i
(nicotinate mononucleotide pyrophosphorylase)
ribose-P
Niacin ribonucleotide
tetrahydro-biopterin
O_2
NAD(P)H + H^+
(tryptophan hydroxylase)
(See Figure 12.16 for details)
$NAD(P)^+$
H_2O
5-Hydroxytryptophan
(tryptophan decarboxylase)
CO_2
2-(1-Acetyl-5-hydroxyindolyl) ethylamine
Serotonin (5-hydroxytryptamine)
5-Hydroxyindoleacetylglycine
[glycine]
H_2O
(monoamine oxidase; MAO)
NH_3
(aldehyde dehydrogenase)
NAD^+
NADH + H^+
5-Hydroxyindolealdehyde
5-Hydroxyindoleacetic acid

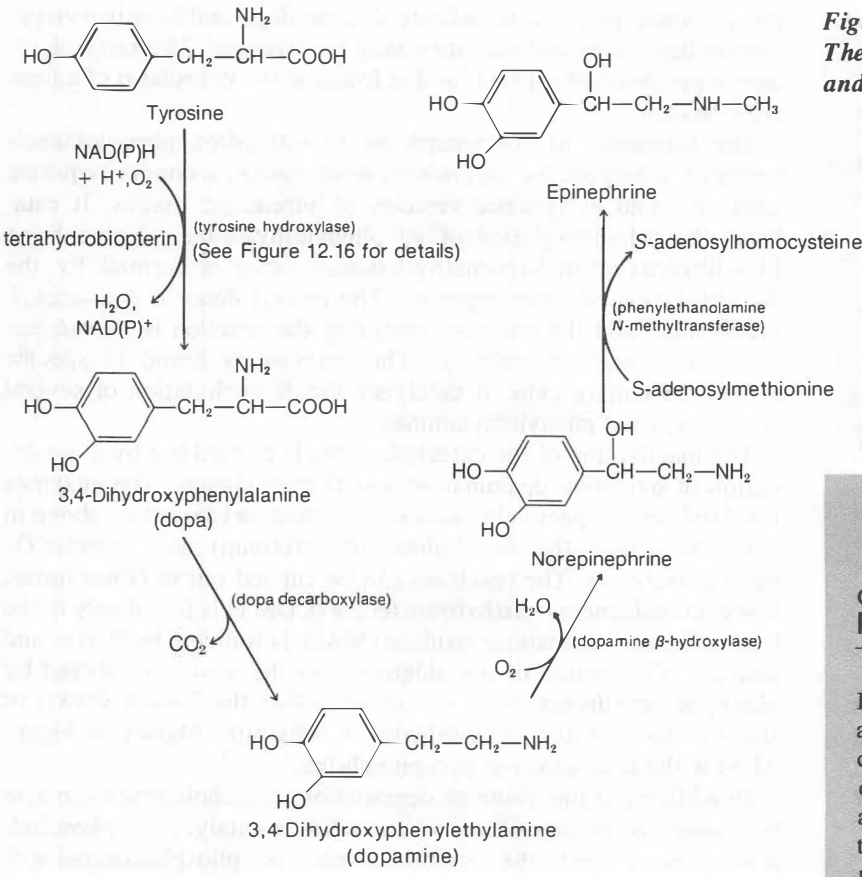


Figure 12.33
The formation of dopamine, norepinephrine, and epinephrine from tyrosine.

dihydroxyphenylalanine). The enzyme that catalyzes this reaction, *tyrosine hydroxylase*, is essentially identical in its action and cofactor requirements to phenylalanine hydroxylase (see above). The enzyme, which is soluble, is found in the central nervous system, in sympathetic ganglia, and in the adrenals. The next step, the decarboxylation of dopa to *dopamine* (*3,4-dihydroxyphenylethylamine*), is catalyzed by *dopa decarboxylase*, an enzyme of rather wide specificity. In addition to dopa, the enzyme will catalyze the decarboxylation of tyrosine, histidine, α-methyldopa, *m*-tyrosine, and possibly tryptophan and 5-hydroxytryptophan. However, as noted above,

Figure 12.32
The formation of serotonin and niacin ribonucleotide from tryptophan.

CLIN. CORR. 12.12 PARKINSON'S DISEASE

Parkinson's disease is the only disorder of amino acid metabolism considered in this chapter that is not of genetic origin. It is a disease that generally develops late in life, around the eighth or ninth decade, although earlier development is not uncommon. Its alternate name, *paralysis agitans*, epitomizes the combination of difficult locomotion, flat facial expression, rigidity, and tremor.

Motor activity is mediated through the part of the midbrain called the *substantia nigra*, so named because of the heavy deposits of melanin it contains. In Parkinson's disease, these areas (and others) become depigmented and lose melanocytes. There is a loss of dopamine, norepinephrine, and serotonin; the loss of dopamine is specific for the putamen, the caudate nucleus, and most important, the substantia nigra, where it is the principal neurotransmitter. It has been suggested that the loss of norepinephrine and serotonin is associated with the depression that some-

times accompanies the disease. The concomitant loss of melanin and dopamine from the substantia nigra reflect the common embryological origin of the central nervous system and the melanocytes, and suggest that in the central nervous system the neurotransmitters and melanin may share a common pool of dihydroxyphenylalanine.

The loss of dopamine is correlated with deficiencies in tyrosine hydroxylase, dopa decarboxylase, the uptake system for dopamine, and a drop in the level of homovanillic acid, the principal metabolite (in human beings) of dopamine. The 24-h urinary excretion of norepinephrine, dopa, dopamine, and 5-hydroxyindoleacetic acid (the principal metabolite of serotonin) is depressed; however, the excretion of epinephrine, primarily associated with peripheral tissues, is normal.

The connection between Parkinson's disease and lowered catecholamine levels is further supported by the phenomenon of drug-induced Parkinsonism. Drugs that lower catecholamine levels (reserpine, α-methyl-*p*-tyrosine) and drugs that block catecholamine receptors (antipsychotic drugs of the phenathiazine and butyrophenone series) produce symptoms that can result in a mistaken diagnosis of Parkinson's disease.

The underlying pathology is the destruction of the nigrostriatal dopaminergic pathway. This explains the various neuropathological and biochemical findings which, in turn, provide a basis for rational therapy. Since the early 1960s, this has consisted in administration of L-dopa (levodopa); more recently this has been supplemented with carbidopa [α-methyl-α-hydrazino-β-(3,4-dihydroxyphenyl)propionic acid]. L-Dopa is the precursor of dopamine (Figure 12.33), which does not cross the blood–brain barrier: L-dopa does, so that a sufficiently high blood level can be provided for the brain

there is some evidence to indicate that the dopa and 5-hydroxytryptophan decarboxylase activities may be separate. The enzyme requires pyridoxal phosphate, and is found in the cytoplasm of adrenergic tissues.

The formation of norepinephrine (3,4-dihydroxyphenylethanolamine) is catalyzed by *dopamine-β-hydroxylase,* a copper-requiring enzyme found in synaptic vesicles of adrenergic tissues. It catalyzes the β-hydroxylation of all phenylethylamines. Epinephrine (3,4-dihydroxyphenyl-(*N*-methyl)ethanolamine) is formed by the *N*-methylation of norepinephrine. The methyl donor is *S*-adenosylmethionine, and the enzyme catalyzing the reaction is *phenylethanolamine* N-*methyltransferase.* This enzyme is found in specific adrenal medullary cells; it catalyzes the *N*-methylation of several β-hydroxylated phenylethylamines.

The inactivation of the catecholamines is carried out by a combination of oxidative deamination and *O*-methylation. The enzymes involved are, respectively, monoamine oxidase (described above in connection with the metabolism of serotonin) and *catechol-O-methyltransferase.* The reactions can be carried out in either order; however, catechol-*O*-methyltransferase (COMT) is found only in the liver, whereas monoamine oxidase (MAO) is found in both liver and neurons. Conversion of the aldehydes to the acid is catalyzed by aldehyde dehydrogenase; it can utilize either the 3,4-dihydroxy- or the 3-methoxy-4-hydroxyaldehyde as substrate. Shown in Figure 12.34 is the catabolism of norepinephrine.

In addition to this route of degradation, catecholamines can also be conjugated as the sulfates. The reaction is catalyzed by *phenolsulfotransferase,* and the sulfate donor is phosphoadenosine-5-phosphosulfate (page 522). In liver, the sulfur for this compound arises in the degradation of cysteine (page 590); the synthetic pathway is not well worked out in this tissue, but in yeast it involves the formation of adenosine phosphosulfate, followed by phosphorylation at the 3 position.

Melanin is formed from tyrosine via L-dopa as intermediate; however, the enzyme that catalyzes the formation of L-dopa in this pathway is not tyrosine hydroxylase, but a copper-containing oxygenase known as *tyrosinase.* It was first isolated from mushrooms, and has been demonstrated in a number of tissues, especially skin. It is also found in some of the same tissues as tyrosine hydroxylase, such as the substantia nigra of the brainstem. This is more than an accidental circumstance; the melanocytes and the cells which differentiate into the central nervous system have a common embryological origin, and in both phenylketonuria (Clin. Corr. 12.9) and Parkinson's disease (Clin. Corr. 12.12) the substantia nigra loses pigmentation.

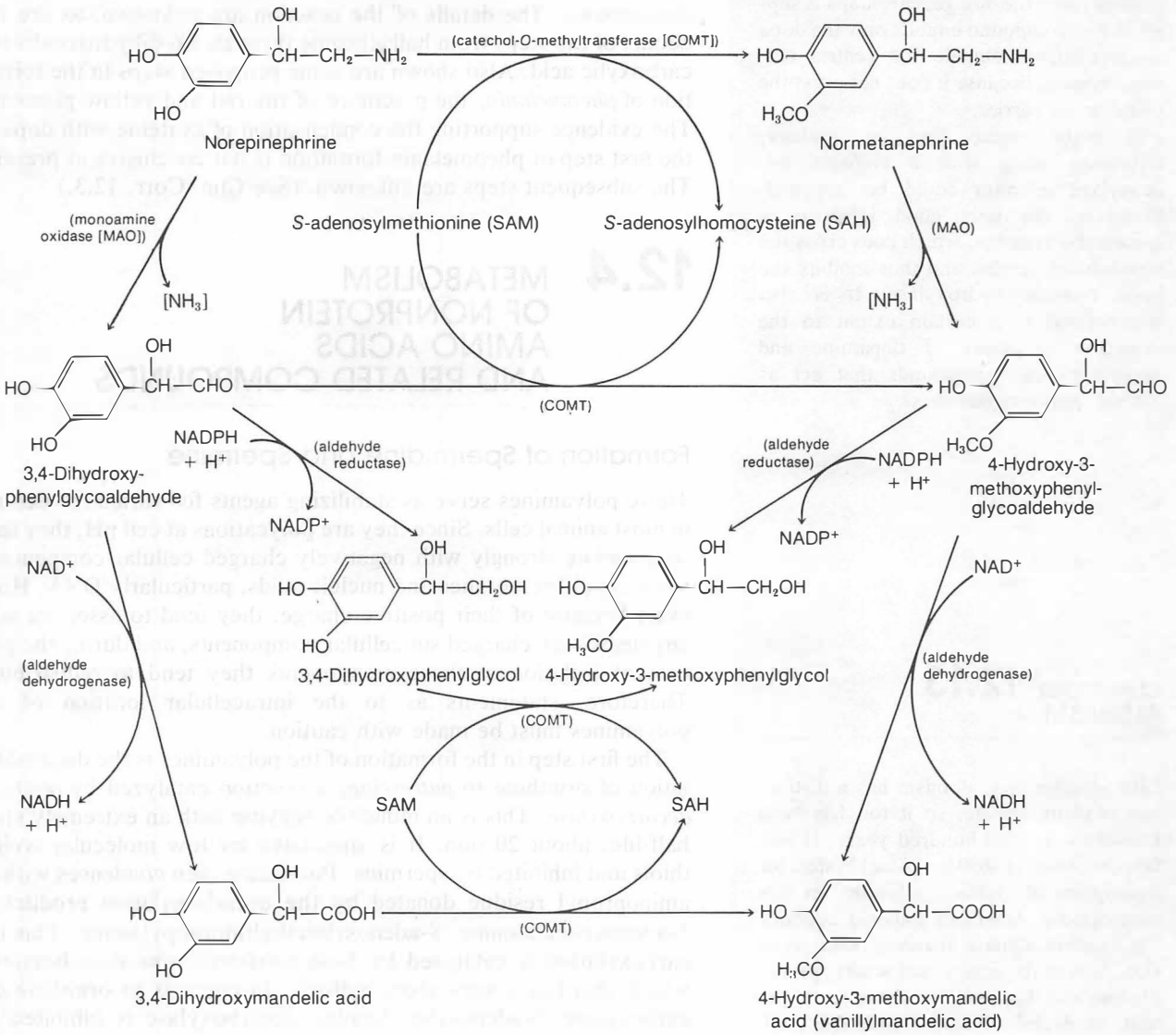


Figure 12.34
The catabolism of norepinephrine.

dopa decarboxylase, despite the relative deficiency of the enzyme, to produce normal or near-normal levels of dopamin. However, administering high levels of L-dopa has serious side effects, so in order to keep down the dosage, carbidopa is supplied: this compound inhibits only the dopa decarboxylases outside the central nervous system, because it does not cross the blood–brain barrier.

It might appear that by analogy, L-tyrosine along with a tyrosine hydroxylase inhibitor could be supplied. However, the only good inhibitor is α-methyl-*p*-tyrosine, which does cross the blood–brain barrier, and thus inhibits the brain tyrosine hydroxylase. It is also metabolized to a certain extent to the α-methyl analogues of dopamine and norepinephrine, compounds that act as "false" neurotransmitters.

CLIN. CORR. 12.13 ALBINISM

Like alcaptonuria, albinism has a distinct and obvious feature, so it too has been known for several hundred years. It was first described in 1699 by Lionel Wafer; his description of classical albinism (in this case, among American Indians) contains the essential clinical features: milk-white skin, lack of daytime visual acuity, photophobia, and the recessive character of the trait. Sir Archibald Garrod, who also studied alcaptonuria (Clin. Corr. 12.10), was the first to propose that albinism was an

The steps leading to the formation of *eumelanin*, the precursor of the brown and black pigments, are shown in Figure 12.35. It is thought that some steps are nonenzymatic, and that tyrosinase can catalyze all the enzymatic steps involved. *Hallachrome*, the precursor of *5,6-dihydroxyindole-2-carboxylic acid*, is generated by dismutation of a molecule of the leuko compound with a molecule of *dopaquinone*. The details of the reaction are unknown, as are the details of the steps from hallachrome through 5,6-dihydroxindole-2-carboxylic acid. Also shown are some proposed steps in the formation of *pheomelanin*, the precursor of the red and yellow pigments. The evidence supporting the condensation of cysteine with dopa in the first step of pheomelanin formation is not conclusive at present. The subsequent steps are unknown. (See Clin. Corr. 12.3.)

12.4 METABOLISM OF NONPROTEIN AMINO ACIDS AND RELATED COMPOUNDS

Formation of Spermidine and Spermine

These polyamines serve as stabilizing agents for various structures in most animal cells. Since they are polycations at cell pH, they tend to associate strongly with negatively charged cellular components, such as cell membranes and nucleic acids, particularly DNA. However, because of their positive charge, they tend to associate with any negatively charged subcellular components, and during the process of isolation of these components they tend to redistribute. Therefore, statements as to the intracellular location of the polyamines must be made with caution.

The first step in the formation of the polyamines is the decarboxylation of ornithine to *putrescine*, a reaction catalyzed by *ornithine decarboxylase*. This is an inducible enzyme with an extremely short half-life, about 20 min. It is stimulated by low molecular weight thiols and inhibited by spermine. Putrescine then condenses with an aminopropyl residue donated by the decarboxylation product of *S*-adenosylmethionine, *S*-adenosylmethylthiopropylamine. This decarboxylation is catalyzed by *S-adenosylmethionine decarboxylase*, which also has a very short half-life. In contrast to ornithine decarboxylase, *S*-adenosylmethionine decarboxylase is inhibited by spermine, but stimulated by spermidine and putrescine.

The condensation reaction is catalyzed by *spermidine synthetase*, an enzyme whose properties are not well understood. Spermine is

formed in a similar condensation of an aminopropyl residue with spermine. This reaction is catalyzed by a separate enzyme, *spermine synthetase*. Again, not much is known about this enzyme either, except that it is distinct from spermidine synthetase. The sequence of reactions is shown in Figure 12.36.

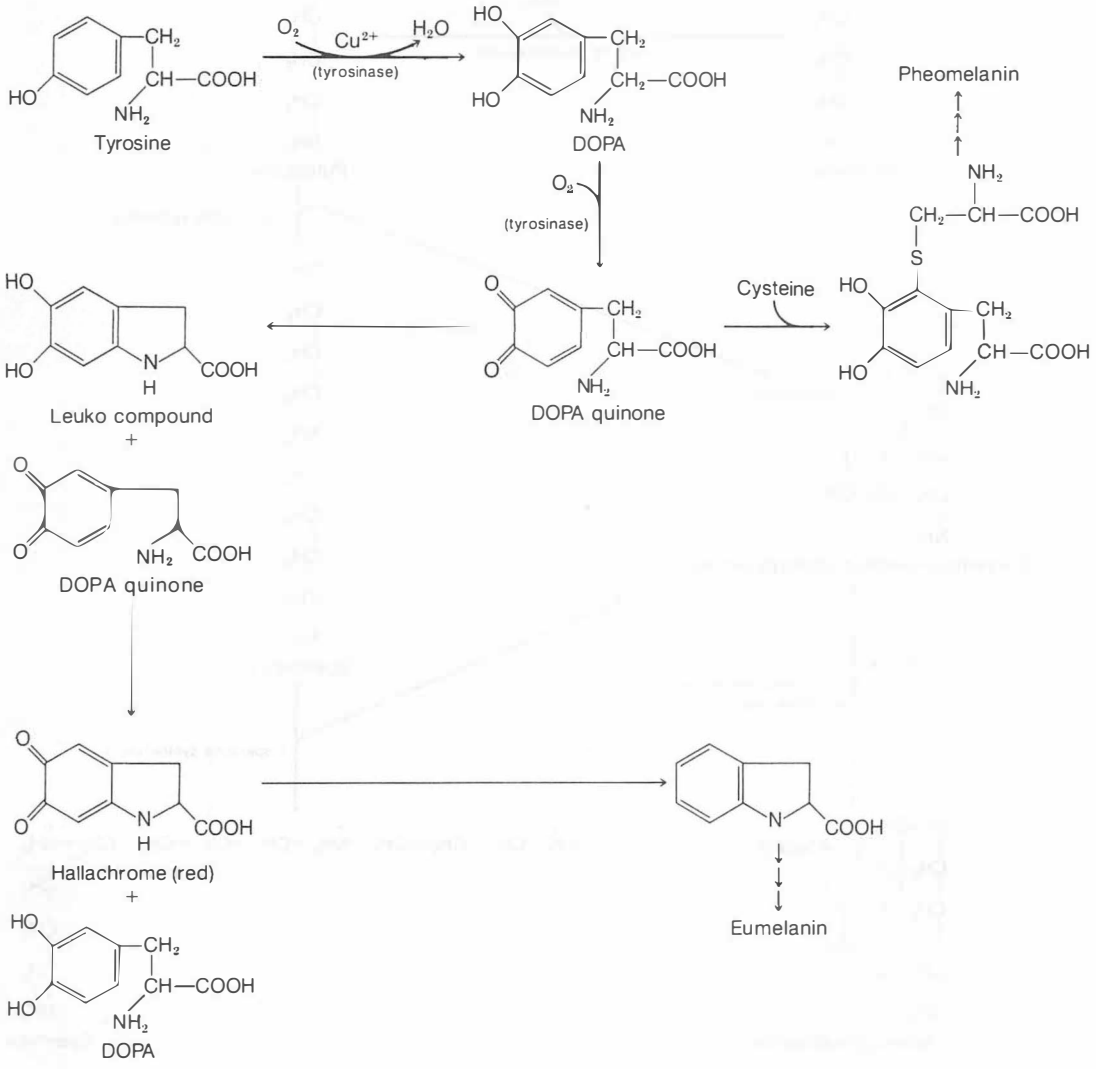


Figure 12.35
The formation of eumelanin and pheomelanin from tyrosine.

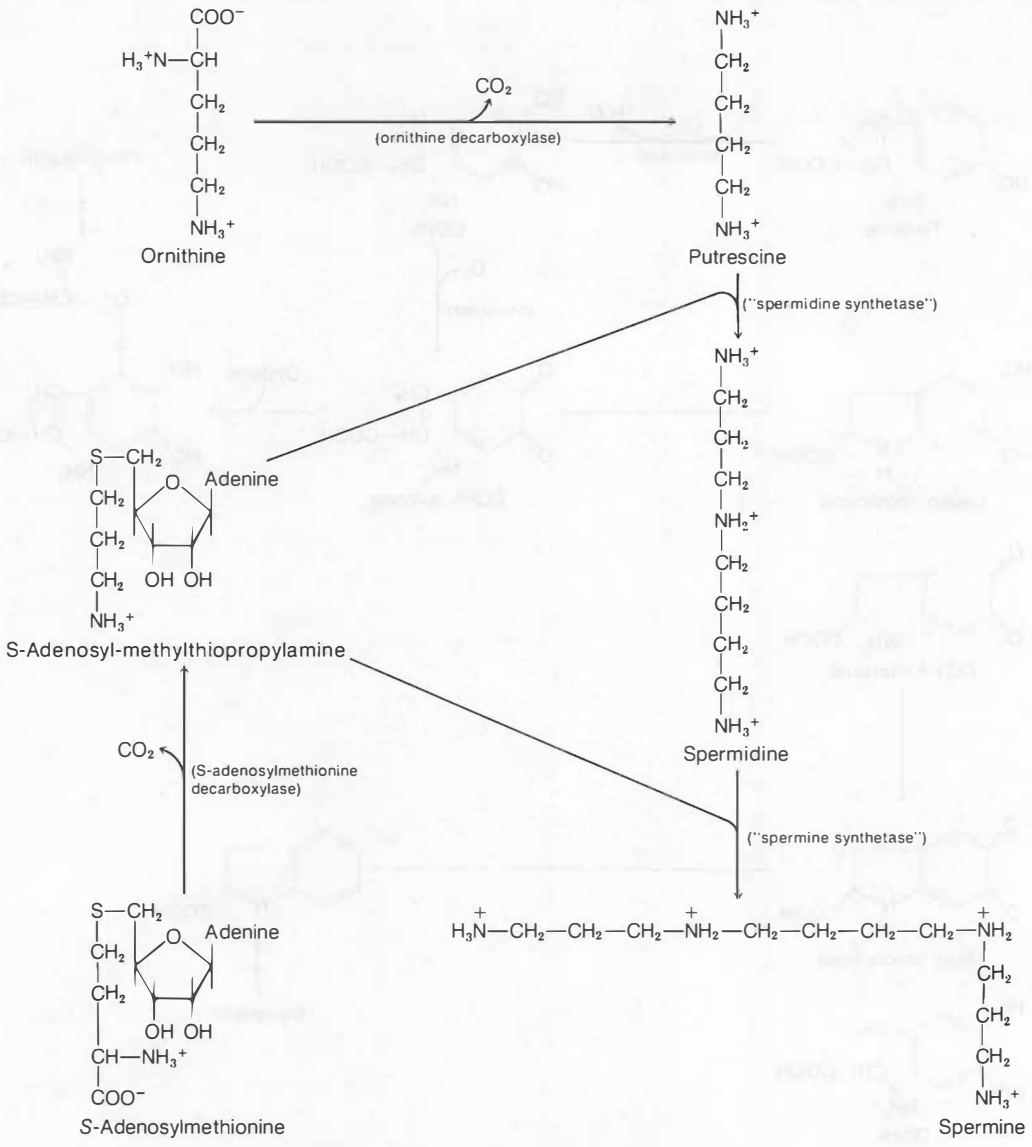
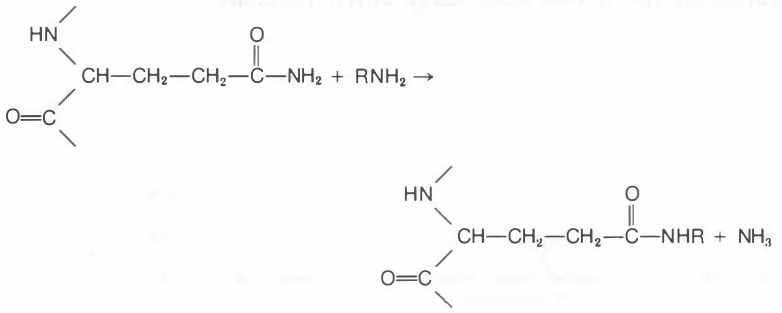


Figure 12.36
The biosynthesis of spermidine and spermine.

A number of acetylated and/or oxidized metabolites of the polyamines have been identified. Among these is *γ-aminobutyric acid* (*GABA*), which is a neurotransmitter. However, this does not seem to be a major pathway for the formation of GABA.

It has been known for some time that polyamines can be covalently linked to proteins. These reactions are catalyzed by *transglutaminases*, a family of enzymes of wide distribution, but which have been demonstrated only in cell-free systems. The acceptor in these reactions is glutamine bound in peptide linkage in a polypeptide chain. The reaction, which proceeds in two steps, has the net effect of replacing the amide-NH_2 with —NHR, where R is the residue of the amino donor. The overall reaction is



The only reaction of this type known to have physiological significance is the formation of ε-(γ-glutamyl)lysine cross-bridges between polypeptide chains. In this case, the lysyl residue on an adjacent polypeptide chain is the amino donor. These linkages are observed in certain hair and skin proteins and in blood clotting. They are also responsible for the formation of the postcoital vaginal plug observed in rats and guinea pigs. Polyamines inhibit this process by competing with the lysyl residue; the physiological significance of this competition is unknown.

Formation of Carnitine

This compound is the carrier for fatty acids into the mitochondria prior to oxidation (Chapter 9). Its synthesis begins with the formation of γ-aminobutyric acid (GABA) from glutamate, a reaction catalyzed by *glutamate decarboxylase;* this is the major route of GABA synthesis. In addition to being a precursor of carnitine, GABA is also a neurotransmitter, as noted above. The next step is three successive methylations with *S*-adenosylmethionine as donor as in the formation of phosphatidylcholine from phosphatidylethanolamine (Figure 12.12); the produce is *γ-butyrobetaine*.

inborn error of metabolism, and indeed suggested at that time (1908) that the defect was a lack or deficiency of melanin formation.

Since then, much information has been obtained concerning the process of melanin formation and normal pigmentation. In addition to the actual synthesis of melanin itself, there are also processes involving the formation and transport of the melanin-containing melanosomes, and defects in these processes will produce variants of classical albinism in which tyrosinase is present. A test for the presence of tyrosinase has been developed, in which hair bulbs are incubated with tyrosine for 12 h and then examined under light microscopy for the presence of pigment. Using this test, a number of different types of albinism, almost all of which involve pigmentation defects of both skin and eyes, have been described. The only tyrosinase-negative (ty-neg) type is the classical albinism described by Wafer; electron microscope examination of hair bulbs from affected individuals shows that the melanosomes are essentially undeveloped.

Tyrosinase-positive (ty-pos) albinos have some pigment in hair bulbs. At birth, they look like ty-neg albinos, but as they grow older some pigmentation develops. Also, in contrast to ty-neg albinos, whose visual problems worsen with age, the visual acuity of ty-pos albinos seems to increase.

Two other types of ty-pos albinos will be mentioned here; the Hermansky–Pudlak syndrome (HPS) and the Chédiak–Higashi syndrome (CHS). In addition to albinism, HPS patients show a defect in platelets in which (among other things) there is a virtual absence of dense bodies; such platelets fail to aggregate under appropriate stimuli. Clinically, frequent episodes of bleeding (a few fatal) are a distinct feature

of this disease. CHS is a fatal childhood disease, characterized by ty-pos albinism and the presence of giant melanosomes unable to migrate properly.

Except for the classical ty-neg albinism, the basic defect in the various syndromes is unknown.

The enzyme system responsible for the formation of this compound has not been isolated. γ-Butyrobetaine is then hydroxylated to form carnitine; this reaction is catalyzed by *γ-butyrobetaine hydroxylase*, a soluble liver enzyme, which requires NADPH and Fe^{2+}. The steps in the synthesis are shown in Figure 12.37.

At this point, it is useful to look back again at the information that has been accumulated and to attempt a summary. Recall that Figure 12.13 showed the oxidative breakdown of the protein amino acids grouped according to end product. Using this as a base, we can incorporate the information regarding the function of the amino acids in various precursor roles. This is shown in Figure 12.38, in which the information from Figure 12.13 is not boxed. Note that relatively few amino acids serve other functions.

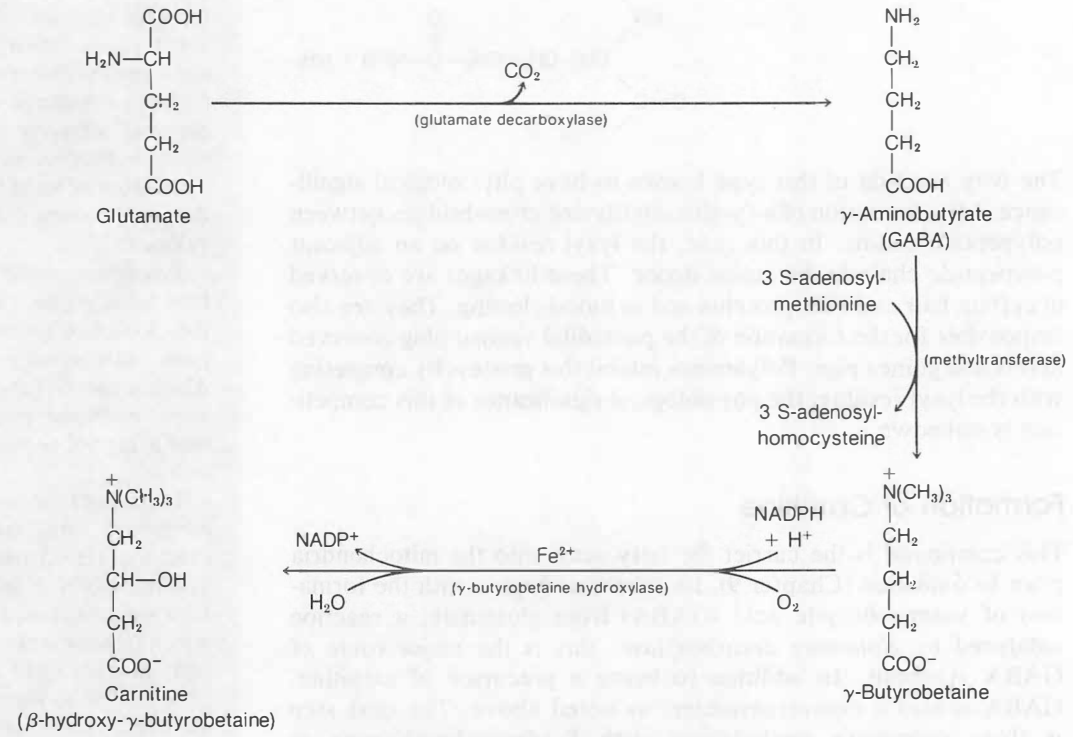


Figure 12.37
The biosynthesis of carnitine.

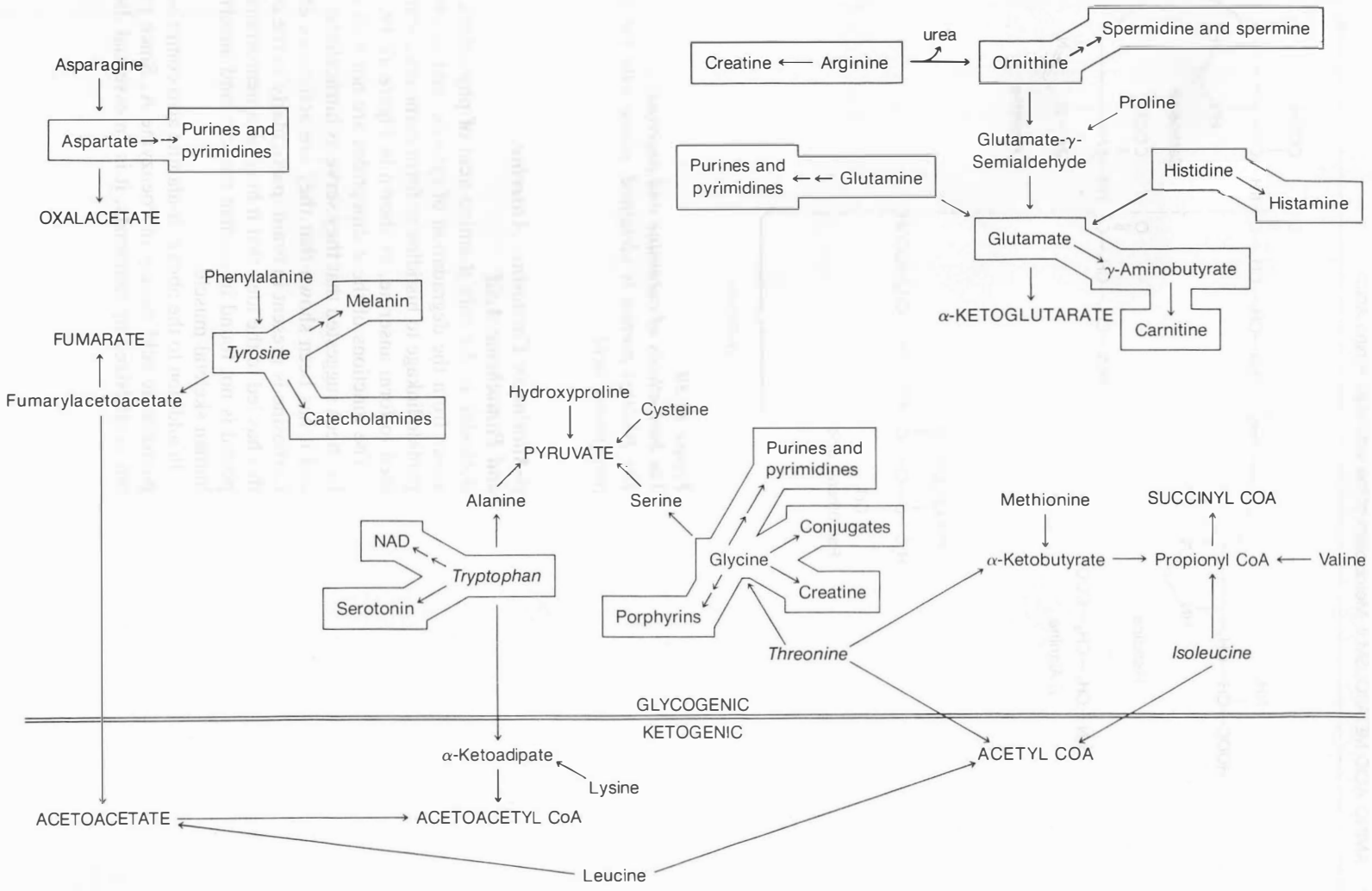


Figure 12.38
Precursor roles of a number of amino acids.

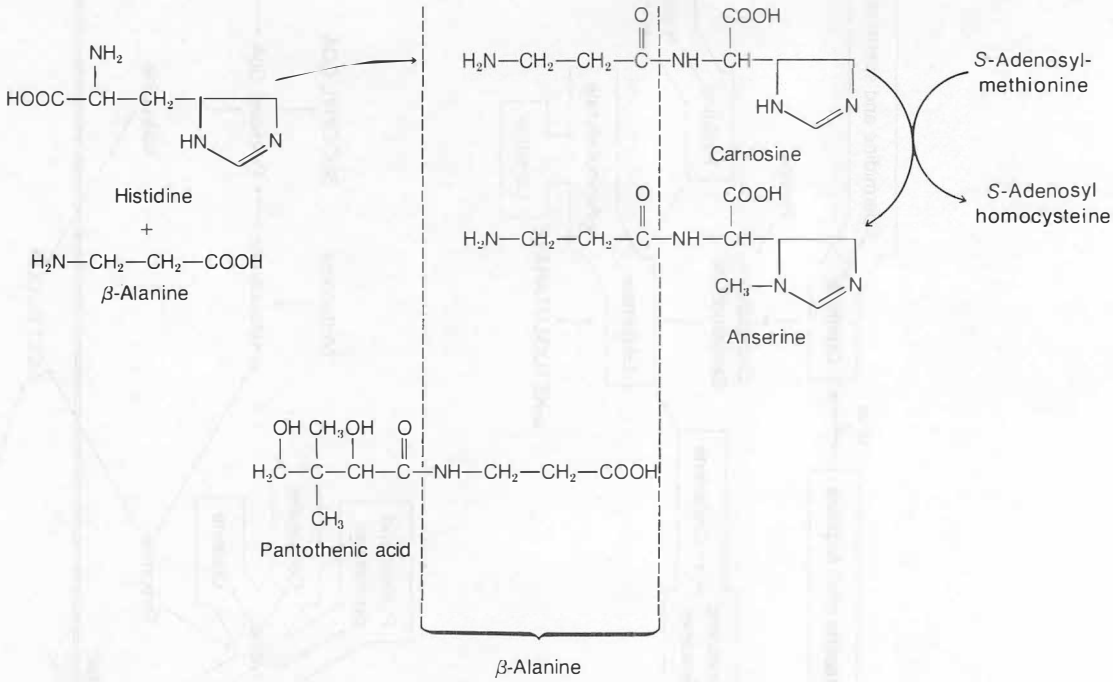


Figure 12.39
The biosynthesis of carnosine and anserine.
The β-alanyl portion is identified, along with the β-alanyl portion of pantothenic acid.

β-Alanine in Carnosine, Anserine, and Pantothenic Acid

β-Alanine is the only β-amino acid of physiological significance. It arises from the degradation of *cytosine* and *anserine;* it is joined in peptide linkage to histidine to form carnosine, which is then methylated to form anserine, as shown in Figure 12.39.

The functions of these dipeptides are not well understood, but it has been suggested that they serve as intracellular buffers in muscle, and it has been shown that they are activators of myosin ATPase. Carnosine is present in brain, particularly in the olfactory pathway; this has led to the idea that it may be a neurotransmitter. This compound is not found in cardiac muscle, and anserine is absent from human skeletal muscle.

In addition to the above, β-alanine also comprises a portion of the pantothenic acid moiety of coenzyme A. Since pantothenic acid is not synthesized by mammals, it is an essential dietary component.

The structures of carnosine, anserine, and pantothenic acid, with the β-alanyl portion identified, are shown in Figure 12.39.

12.5 AMINO ACID TRANSPORT

According to an interesting hypothesis proposed by Meister, the tripeptide *glutathione* (*γ-glutamylcysteinylglycine*) participates in the transport mechanism which also involves the participation of six known enzymes, one of which is membrane-bound, the other five being in the cytosol. The hypothesis provides a partial answer to the question of amino acid transport across cell membranes.

The proposed mechanism has been called the *γ-glutamyl cycle* (Figure 12.40). The cycle starts with the displacement of *cys-*

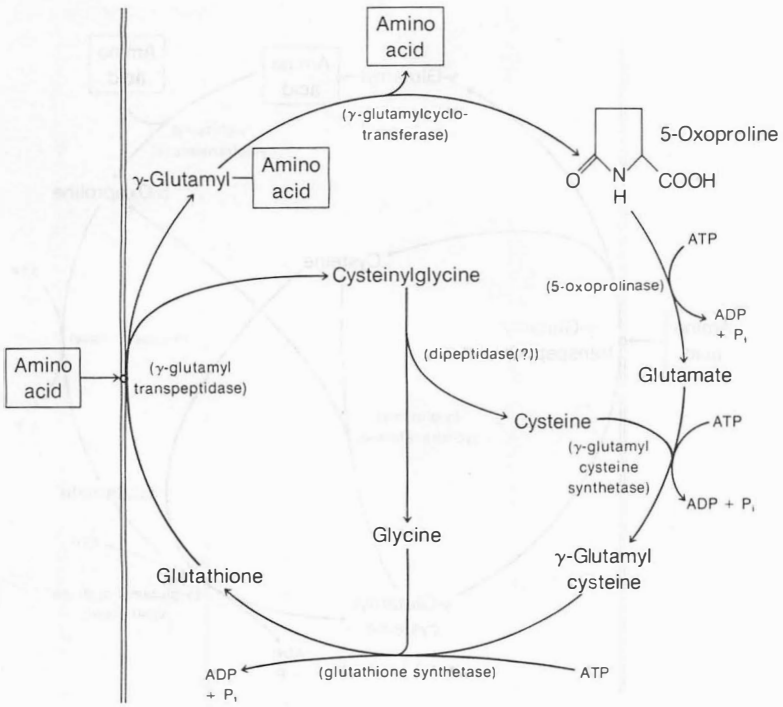


Figure 12.40
The γ-glutamyl cycle.

CLIN. CORR. 12.14 5-OXOPROLINURIA

Oxoprolinuria is characterized by a severe metabolic acidosis that, if untreated, leads to pronounced mental and developmental deficiencies. The acidosis is caused by excessive concentration of 5-oxoproline, an intermediate in the γ-glutamyl cycle. Originally, it was thought that the deficient enzyme was 5-oxoprolinase, which catalyzes the conversion of 5-oxoproline to glutamate. It has now been shown, however, that the defect is in glutathione synthetase.

Glutathione is an inhibitor of γ-glutamylcysteine synthetase. When the level of glutathione is reduced, there is an increased synthesis of γ-glutamylcysteine. This dipeptide, in addition to being a substrate for glutathione synthetase, is also a good substrate for γ-glutamyltranspeptidase and γ-glutamylcyclotransferase; it can be broken down in the presence of either of these two enzymes to yield cysteine and 5-oxoproline. This modified γ-glutamyl cycle is shown in Figure 12.41. Under these conditions, there is generated an excess of 5-oxoproline. A variant of this disorder has been described in which the glutathione synthetase deficiency is restricted to erythrocytes. The affected individuals do not have 5-oxoprolinuria and are clinically well.

Two patients have been reported with a deficiency of γ-glutamylcysteine synthetase. These patients (brother and sister) had a generalized aminoaciduria, hemolytic anemia, and muscle and nervous system abnormalities, which did not develop until young adulthood. With this enzyme deficit, glutathione levels are again reduced, but no alternative cycle can operate, owing to the shortage of γ-glutamylcysteine.

teinylglycine from glutathione and its replacement by the incoming amino acid; this reaction is catalyzed by the membrane-bound *γ-glutamyl transpeptidase.* The resulting cysteinylglycine and γ-glutamyl-amino acid are now in the cytoplasm. The amino acid is derived from the glutamate in a reaction catalyzed by *γ-glutamylcyclotransferase;* the other product is *5-oxoproline.* The cysteinylglycine is probably cleaved by *cysteinylglycine dipeptidase.*

At this point we have, in addition to the free amino acid, 5-oxoproline, cysteine, and glycine. In order for the reaction to operate as a cycle, glutathione must be reconstituted from its amino acids.

The process starts with the regeneration of glutamate from 5-oxoproline. This conversion is catalyzed by *5-oxoprolinase;* the reaction requires ATP and two molecules of water. The glutamate so formed reacts with cysteine to form the dipeptide γ-glutamylcysteine; the reaction is catalyzed by *γ-glutamylcysteine synthetase*

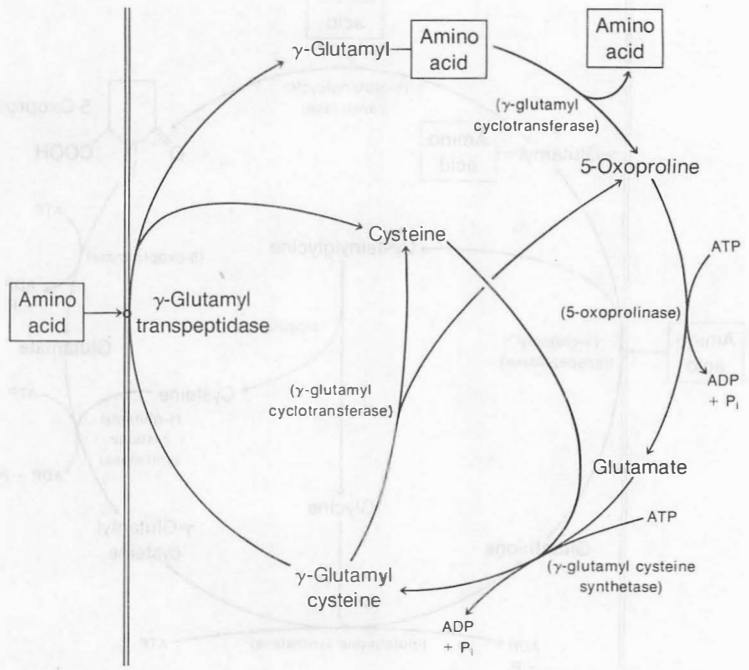


Figure 12.41
The γ-glutamyl cycle modified by the absence of glutathione synthetase.

and is inhibited by glutathione. In the last step, the γ-glutamylcysteine condenses with glycine in a reaction catalyzed by *glutathionine synthetase;* this reaction completes the resynthesis of glutathione. As can be seen from the figure, the net effect of the operation of the cycle is to transport one molecule of an amino acid across the cell membrane at the expense of the hydrolysis of three molecules of ATP. It should be noted that γ-glutamylcysteine is a good substrate for both the membrane-bound transpeptidase and the cyclotransferase, as well as glutathione synthetase. This fact is significant in the understanding of the hereditary disorder 5-oxoprolinuria (Clin. Corr. 12.14).

The γ-glutamyl cycle is the only general proposal put forth to date describing amino acid transport. Nevertheless, it does not cover all situations: it does not include the imino acids (proline and hydroxyproline), and it probably does not operate for acidic amino acids. Furthermore, it may not operate in all tissues. The cycle requires the consumption of three molecules of ATP per amino acid molecule transported, which means that it is an active process; yet amino acid transport takes place in red blood cells, where there is no ATP production.

Clearly, other mechanisms must operate, but at present, these are poorly understood. The possible role of pyridoxal phosphate in amino acid transport is mentioned in Chapter 11. For a recent assessment of the status of amino acid transport, see Lerner (1978).

BIBLIOGRAPHY

General

Boyer, P. *The enzymes,* 3rd ed. Volumes 4–9 and 11–13. New York: Academic, 1970.

Felig, P. Amino acid metabolism in man. *Ann. Rev. Biochem.* 44:933, 1975.

Friedhoff, A. J. *Catecholamines and behavior.* New York: Plenum, 1975.

Meister, A. *Biochemistry of the amino acids,* 2nd ed. New York: Academic, 1965.

Urea Cycle

Grisolia, S., Baguena, R., and Mayor, F. *The urea cycle.* New York: Wiley, 1976.

Holmes, F. L. Hans Krebs and the discovery of the ornithine cycle. *Fed. Proc.* 39:216, 1980.

Polyamines

Williams-Ashman, H. G., and Canellakis, Z. N. Polyamines in mammalian biology and medicine. *Perspect. Biol. Med.* 22:421, 1979.

Transport

Lerner, J. *A review of amino acid transport processes in animal cells and tissues.* Orono: University of Maine at Orono Press, 1978.

Disorders of Amino Acid Metabolism

Conn, H. O., and Lieberthal, M. M. *The hepatic coma syndromes and lactulose.* Baltimore: Williams & Wilkins, 1979.

Rosenberg, L. E., and Scriver, C. R. Disorders of amino acid metabolism, in P. K. Bonay and L. E. Rosenberg (eds.), *Metabolic control and disease,* 8th ed. Philadelphia: W. B. Saunders, 1980, p. 583.

Stanbury, J. B., Wyngaarden, J. B., and Frederickson, D. O. *The metabolic basis of inherited disease,* 4th ed. New York: McGraw-Hill, 1978.

13

Purine and Pyrimidine Nucleotide Metabolism

JOSEPH G. CORY

13.1 OVERVIEW

Purine and pyrimidine nucleotides are critically important metabolites that participate in many cellular functions. These functions range from serving as the monomeric precursors of the nucleic acids, to energy stores, to effectors, to group transfer agents, and to mediators of hormone action. The nucleotides are formed in the cell de novo from amino acids, ribose, formate, and CO_2. The de novo pathway for the synthesis of the nucleotides requires a relatively high input of energy. To compensate for this, most cells have very efficient "salvage" pathways by which the preformed purine or pyrimidine base can be reutilized.

Because of the manner in which nucleotides are synthesized or "salvaged," the purines and pyrimidines occur primarily as nucleotides in the cell. The concentrations of free bases or free nucleosides under normal conditions are exceedingly small. The levels of nucleotides in the cell are very finely regulated by a series of allosterically controlled enzymes in the pathway. Nucleotides are the regulators of these reactions.

Ribonucleotides (at the diphosphate level) serve as the precursors of the deoxyribonucleotides. While the concentrations of ribonucleotides in the cell are in the millimolar range, the concentrations of deoxyribonucleotides are in the micromolar range. DNA replication requires that there be sufficient quantities of the deoxyribonucleoside triphosphates. To facilitate this, the levels of several of the enzymes involved with deoxyribonucleotide metabolism increase just prior to the replication of DNA in the cell.

There are several diseases or syndromes that result from defects in the metabolic pathways for the synthesis of nucleotides either de novo or by salvage or for the degradation of the nucleotides. These include gout, the Lesch–Nyhan syndrome, orotic aciduria, and immunodeficiency diseases. Since nucleotides are obligatory for DNA and RNA synthesis in dividing cells, the metabolic pathways involving the synthesis of nucleotides have been the sites at which many antitumor agents have been directed.

It should be kept in mind during the reading of this chapter on the metabolism of purine and pyrimidine nucleotides that the metabolism discussed has been limited exclusively to mammalian cells. In certain instances there are major differences between the bacterial and mammalian cells in nucleotide synthesis, degradation, and regulation.

13.2 METABOLIC FUNCTIONS OF NUCLEOTIDES

All types of cells (mammalian, bacterial, and plant) contain a wide variety of nucleotides and their derivatives. Some of these nucleotides occur in relatively high concentrations (millimolar range) in the cells. The reason for the large number of nucleotides and their derivatives in the cell is that they are involved in many metabolic processes that must be carried out for normal cellular growth and function.

These functions include the following:

1. *Role in Energy Metabolism:* As we have already seen, ATP is the main form of chemical energy available to the cell. Quantitatively, ATP is generated in cells by oxidative phosphorylation and substrate-level phosphorylation. ATP is utilized to drive metabolic reactions, as a phosphorylating agent, and is involved in such processes as muscle contraction, active transport, and maintenance of cell membrane integrity. As a phosphorylating agent, ATP serves as the phosphate donor for the generation of the other nucleoside 5′-triphosphates (e.g., GTP, UTP, CTP).

2. *Monomeric Units of Nucleic Acids:* The nucleic acids, DNA and RNA, are composed of monomeric units of the nucleotides. In the reactions in which the nucleic acids are synthesized, the nucleoside 5′-triphosphates are the substrates and are linked in the polymer through 3′,5′-phosphodiester bonds with the release of pyrophosphate.

3. *Physiological Mediators:* More recently recognized functions of nucleotides and their derivatives involve those in which the nucleotides or nucleosides serve as mediators of key metabolic processes.

The role of cAMP as a "second messenger" in epinephrine- and glucagon-mediated control of glycogenolysis and glycogenesis has already been discussed. The importance of cGMP as a mediator of cellular events has also been recognized.

ADP has been shown to be very critical for normal platelet aggregation and hence blood coagulation.

Adenosine has been shown to cause dilation of coronary blood vessels and therefore may be important in the regulation of coronary blood flow.

4. *Components of Coenzymes:* Coenzymes such as NAD, FAD, and coenzyme A are important metabolic constituents of cells and are involved in many metabolic pathways.

NAD is a coenzyme (cosubstrate) that is involved in oxidation–reduction reactions. This coenzyme contains AMP as part of the molecule. While the AMP moiety is not directly involved in the electron transfer, it is critical for the binding of the coenzyme to the particular enzyme.

FAD likewise is a coenzyme involved in electron transfer reactions. AMP is also part of this molecule and serves the same function as it does in NAD, that is, the binding of the coenzyme to the enzyme.

Coenzyme A functions as an acyl group transfer agent (e.g., acetyl CoA, palmityl CoA). In this coenzyme, the nucleotide present is adenosine 3′,5′-diphosphate. The 3′,5′-ADP moiety is not involved in binding the acyl groups (which are actually bound as a thiol ester), but appears to be critical for the binding of the coenzyme to the appropriate enzyme.

5. *Activated Intermediates:* The nucleotides also serve as carriers of "activated" intermediates required for a variety of reactions. A compound such as UDP-glucose is a key intermediate in the synthesis of glycogen and glycoproteins. GDP-mannose, GDP-fucose, UDP-galactose, and CMP-sialic acid are all key intermediates in reactions in which sugar moieties are transferred for the synthesis of glycoproteins. CTP is utilized to generate CDP-choline, CDP-ethanolamine, and CDP-diglycerides, which are involved in phospholipid metabolism.

6. *Allosteric Effectors:* Many of the regulated steps of the metabolic pathways are controlled by the intracellular concentrations of nucleotides. A few examples of these are listed in Table 13.1.

These few examples give an indication of the diverse roles nucleotides and their derivatives play in normal cell metabolism.

Table 13.1 Nucleotides as Allosteric Effectors

Enzyme	*Activated By*	*Inhibited By*
Phosphofructokinase	AMP	ATP
Fructose bisphosphatase		AMP
Isocitrate dehydrogenase	ADP	NADH
CDP reductase	ATP	dATP
dCMP deaminase	dCTP	dTTP

Occurrence in Cells

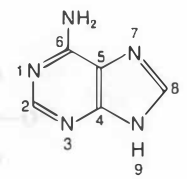
The principal form of purine and pyrimidine compounds found in cells is as the 5′-nucleotide derivative. In normally functioning cells the nucleotide of highest concentration is ATP. Depending on the cell type, the concentrations of the nucleotides vary greatly. For example, in the red cell the adenine nucleotides far exceed the other nucleotides, which are barely detectable. In the liver cells and other tissues a complete profile of the mono-, di-, and triphosphates are found along with UDP-glucose, UDP-glucuronic acid, NAD^+, NADH, and so on. The presence of the free bases, nucleosides or 2′- and 3′-nucleotides in the acid-soluble fraction of the cell represents degradation products of either the endogenous or exogenous nucleotides or nucleic acids. The presence of the so-called minor bases is due to the degradation of nucleic acids.

The ribonucleotide concentration in the cell is in the millimolar range while the concentration of deoxyribonucleotides in the cell is in the micromolar range. As a specific example, the ATP concentration in Ehrlich tumor cells is 3,600 pmol/10^6 cells, while the dATP concentration in these cells is only 4 pmol/10^6 cells. The deoxyribonucleotide levels, however, are subject to major fluctuations during the cell cycle, in contrast to the ribonucleotide levels, which remain relatively constant.

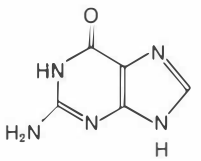
In normal cells the total concentrations of the nucleotides are fixed within rather narrow limits, although the concentration of the individual components can vary. That is, the total concentration of adenine nucleotides (AMP, ADP, ATP) is constant, although there is a variation in the ratio of ATP to AMP + ADP, depending on the energy state of the cells. The basis for this "fixed concentration" is that the synthesis of nucleotides is one of the most finely regulated pathways occurring in the cell, as is discussed later.

13.3 CHEMISTRY OF NUCLEOTIDES

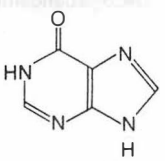
Quantitatively, the major purine derivatives found in the cell are those of adenine and guanine. Other purine bases encountered are hypoxanthine and xanthine. Nucleoside derivatives of these molecules will contain either ribose or 2-deoxyribose linked to the purine ring through a β-*N*-glycosidic bond at N-9. Ribonucleosides contain



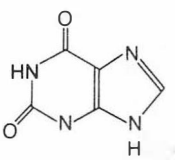
Adenine



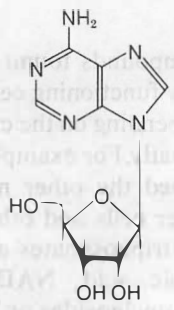
Guanine



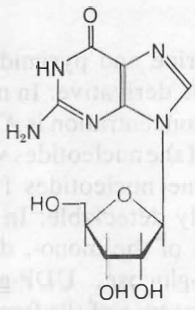
Hypoxanthine



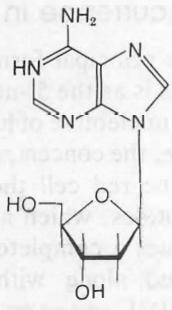
Xanthine



Adenosine

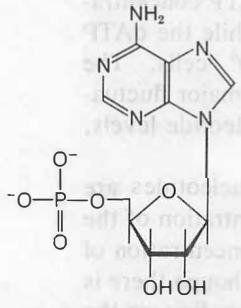


Guanosine

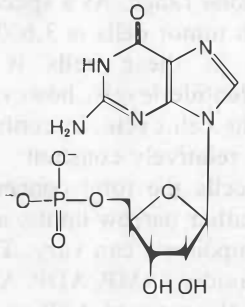


Deoxyadenosine

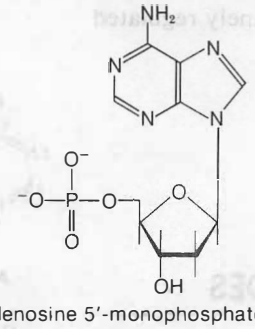
ribose, while deoxyribonucleosides contain deoxyribose as the sugar moiety. Nucleotides are phosphate esters of the purine nucleosides. These are



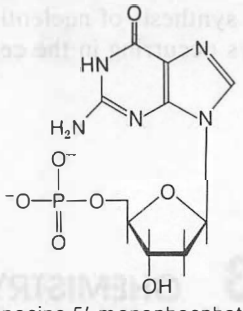
Adenosine 5′-monophosphate (AMP)



Guanosine 5′-monophosphate (GMP)



Deoxyadenosine 5′-monophosphate (dAMP)

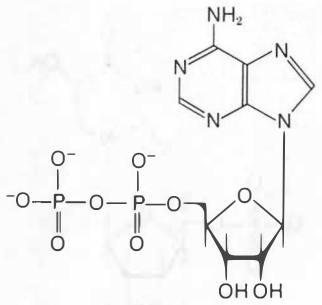


Deoxyguanosine 5′-monophosphate (dGMP)

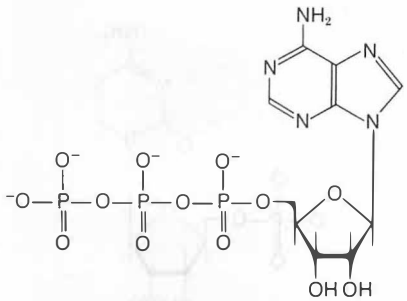
3′-Nucleotides such as adenosine 3′-monophosphate (3′-AMP) can occur in cells as a result of nucleic acid degradation.

In normally functioning cells, the tri- and diphosphates of the nu-

cleosides are found to a greater extent than the monophosphates, nucleosides, or free bases.

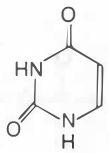


Adenosine 5′-diphosphate (ADP)

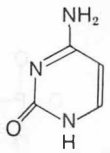


Adenosine 5′-triphosphate (ATP)

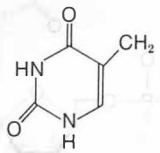
The pyrimidine nucleotides found in highest concentrations in the cell are those containing uracil, cytosine, and thymine. Uracil and



Uracil

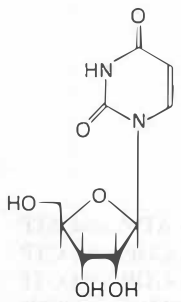


Cytosine

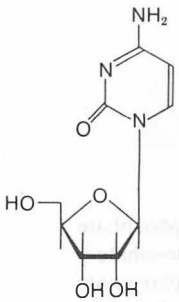


Thymine

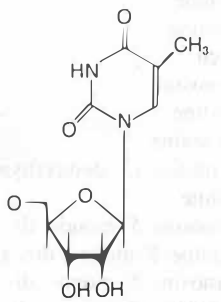
cytosine nucleotides are the major pyrimidine components of RNA, whereas cytosine and thymine are the major pyrimidine components of DNA. As with purine derivatives, the pyrimidine nucleosides or nucleotides contain either ribose or 2-deoxyribose. The sugar moiety is linked to the pyrimidine in a β-N-glycosidic bond at N-1. The nucleosides of the pyrimidines are uridine, cytidine and thymidine.



Uridine

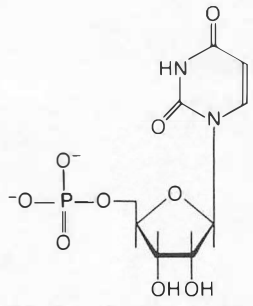


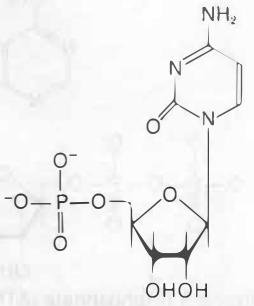
Cytidine

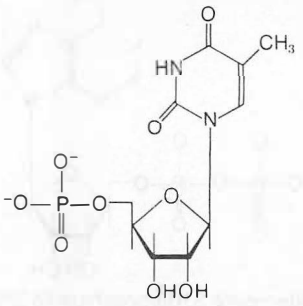


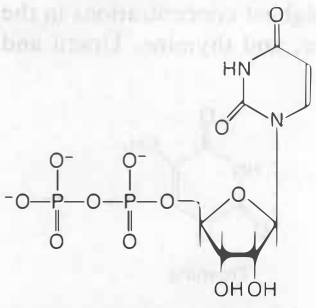
Thymidine

The phosphate esters of the pyrimidine nucleosides are UMP, CMP, and TMP. In the cell the major pyrimidine derivatives found are the tri- and diphosphates.


Uridine 5′-monophosphate (UMP)


Cytidine 5′-monophosphate (CMP)


Thymidine 5′-monophosphate (TMP)


Uridine 5′-diphosphate (UDP)

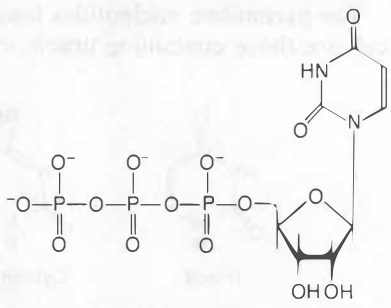

Uridine 5′-triphosphate (UTP)

Table 13.2 Symbols for Bases, Nucleosides, and Nucleotides

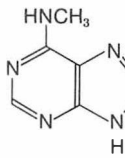
Compound	*Abbreviations*
Adenine	Ade
Cytosine	Cyt
Guanine	Gua
Thymine	Thy
Uracil	Ura
Adenosine	Ado
Cytidine	Cyd
Guanosine	Guo
Thymidine (2′-deoxythymidine)	dThd
Uridine	Urd
Adenosine 5′-mono-, di-, and triphosphate	AMP, ADP, and ATP
Cytidine 5′-mono-, di-, and triphosphate	CMP, CDP, and CTP
Guanosine 5′-mono-, di-, and triphosphate	GMP, GDP, and GTP
Thymidine 5′-mono-, di-, and triphosphate	TMP, TDP, and TTP
Uridine 5′-mono-, di-, and triphosphate	UMP, UDP, and UTP

The symbols and abbreviations for the bases, nucleosides, and nucleotides are summarized in Table 13.2.

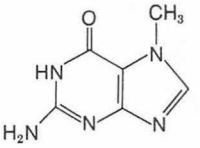
Modified Bases

The modified bases are formed by alteration of the purine or pyrimidine ring only *after* the parent base has been incorporated into the nucleic acids. Examples of some of these modified bases are 6-methylaminopurine, 7-methylguanine, and 5-methylcytosine. The term "minor base" is frequently used and indicates only that these modified bases are found in small quantities relative to adenine, guanine, cytosine, uracil, and thymine. The modified bases found in urine are a direct measure of the turnover of nucleic acids in the cells.

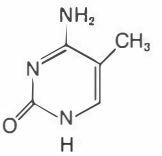
An additional modified nucleoside is also found as a constituent of tRNA. This modified nucleoside is called pseudouridine and is unusual in that this nucleoside contains a C-glycosidic bond rather than a N-glycosidic bond as shown in the structure for pseudouridine. As with the modified bases, pseudouridine is formed only after UMP has been incorporated into the RNA. The level of pseudouridine in the urine is also an excellent measure of tRNA turnover. Ribothymidine is found as a minor component of tRNA. This derivative is formed by the methylation of uracil at C-5 by *S*-adenosylmethionine as a component of tRNA rather than by the direct incorporation of a ribothymidine nucleotide into this RNA species.



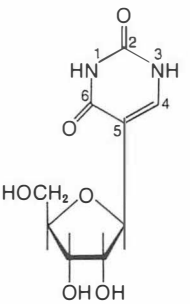
6-Methylaminopurine



7-Methylguanine



5-Methylcytosine



Pseudouridine

Properties of Nucleotides

Cellular components containing either the purine or pyrimidine bases can be easily detected because of the strong absorption of uv light by these compounds. The purine bases, nucleosides and nucleotides have stronger absorptions than the pyrimidines and their derivatives. The molar extinction coefficients (a measure of the light absorption at a specific wavelength of a compound) and λ_{max} for these are given in Table 13.3. The wavelength of light at which maximum absorption occurs varies with the particular base component, but in most cases the uv maximum is close to 260 nm. The uv spectrum for each of the nucleoside or nucleotide derivatives responds differently to changes in pH. The strong uv absorptions and the differences due to the specific structure of the base moiety provide the basis for sensitive methods in assaying these compounds both qualitatively and quantitatively. For example, the deamination of cytosine nucleosides or nucleotides to the corresponding uracil derivatives causes a marked shift in λ_{max} from 271 to 262 nm, which

Table 13.3 Spectrophotometric Constants for Purine and Pyrimidine Nucleosides

Nucleoside	*Molar Extinction Coefficient* $\times 10^{-3}$	λ_{max} *pH 7*
Adenosine	15.4	259
Guanosine	13.7	253
Cytidine	8.9	271
Uridine	10.0	262
Thymidine	10.0	262

is easily determined. Because of the high molar extinction coefficients of the purine and pyrimidine bases and their high concentrations in the nucleic acids, a solution of RNA or DNA at a concentration of 1 mg/ml would have an absorbance at 260 nm of ~20, whereas a typical protein at a concentration of 1 mg/ml would have an absorbance at 280 nm of ~1. Consequently, the nucleic acids are easily detected at low concentrations.

The N-glycosidic bond of the purine and pyrimidine nucleosides and nucleotides are stable to alkali. However, the stability of this bond to acid hydrolysis differs markedly. The N-glycosidic bond of purine nucleosides/nucleotides is easily hydrolyzed by dilute acid at elevated temperatures (e.g., 60°C) to yield the free purine base and the sugar or sugar phosphate. On the other hand, the N-glycosidic bond of uracil, cytosine, and thymine nucleoside/nucleotides is very stable to acid treatment. Strong conditions, such as perchloric acid (60%) and 100°C, will cause the release of the free pyrimidine but with the complete destruction of the sugar moiety. The N-glycosidic bond of the pyrimidine nucleoside/nucleotide containing dihydrouracil is labile to mild acid treatment.

Because of the highly polar phosphate group, the purine and pyrimidine nucleotides are considerably more soluble in aqueous solutions than are their nucleosides and free bases. In general, the nucleosides are more soluble than the free purine or pyrimidine base.

The purine and pyrimidine bases and their nucleoside and nu-

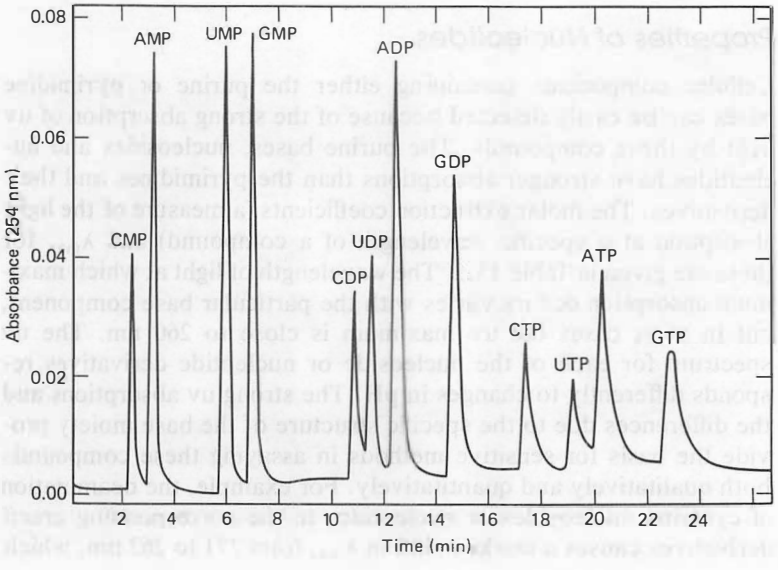


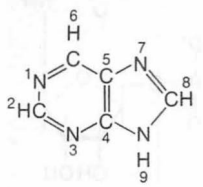
Figure 13.1
Separation of nucleotides by high pressure liquid chromatography.
A mixture of the 12 ribonucleotides was injected onto a Partisil SAX column (25 × 0.46 cm). A linear gradient was developed using 0.01 M ammonium phosphate, pH 2.77 and 0.5 M ammonium phosphate, pH 4.8 at a rate of 5%/min. The solvent flow rate was 2 ml/min.

cleotide derivatives can be easily separated by a variety of techniques. These methods include paper chromatography; thin-layer chromatography, utilizing plates with cellulose or ion-exchange resins; electrophoresis; and ion-exchange column chromatography. The most recent advance in the separation of the purine and pyrimidine components involves the use of high pressure liquid chromatography (HPLC). Commercial instruments are available, which consist essentially of a pump or pumps, an injection port, a column, a uv detector, and a recorder. Using this instrument nanomole quantities of these components are easily separated in a brief period of time. Figure 13.1 shows the separation by HPLC of a mixture containing AMP, ADP, ATP, GMP, GDP, GTP, UMP, UDP, UTP, CMP, CDP, and CTP in 25 min. The development of this instrument and the high resolution columns (anion and cation exchange and reverse phase) has allowed rapid determination of nucleoside and nucleotide pools under a variety of cellular conditions.

13.4 METABOLISM OF PURINE NUCLEOTIDES

The purine ring is synthesized de novo in mammalian cells utilizing amino acids as carbon and nitrogen donors and formate and CO_2 as carbon donors.

The numbering system for the purine ring is shown below with the sources for the various carbon and nitrogen atoms indicated.



Purine

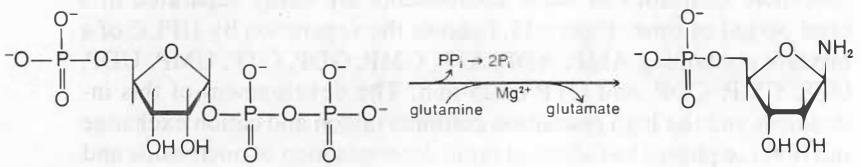
N-3 and N-9 from amide N of glutamine
C-4, C-5, and N-7 from glycine
C-2 and C-8 from formate via H_4folate
C-6 from CO_2
N-1 from aspartate

From the sources of carbon and nitrogen which make up the purine ring it is very evident that amino acids play an important role in nucleotide metabolism. It has been shown that glutamine and aspartate levels can influence the rate of purine nucleotide synthesis in tumor cells. Many of the reactions that are required for the de novo synthesis utilize the hydrolysis of ATP to drive the particular reaction. The overall set of reactions leading to the synthesis of a purine nucleotide is, therefore, expensive in terms of ATP required.

De Novo Synthesis

All the enzymes involved with purine nucleotide synthesis and degradation are found in the cytosol of the cell. However, not all cells are capable of de novo purine nucleotide synthesis. The reactions leading to the de novo synthesis of purine nucleotides are as follows.

1. *Formation of N-Glycosidic Bond* (N-9 of purine ring and C-1 of ribose)

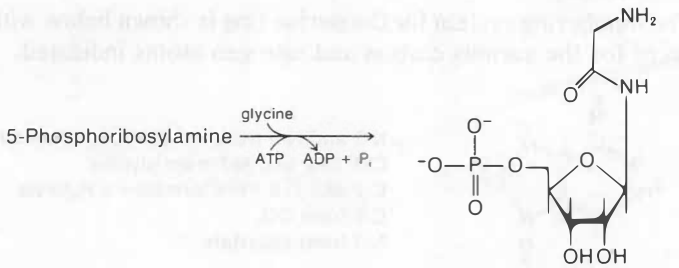


5-Phosphoribosyl pyrophosphate (PRPP) 5-Phosphoribosylamine

This reaction is catalyzed by the enzyme PRPP amidotransferase, which is the committed step in this pathway and, as we will see later, the major regulated step.

In the formation of 5-phosphoribosylamine from PRPP there is inversion at carbon-1, giving rise to the β configuration of the N-glycosidic bond in purine nucleotides.

2. *Addition of Glycine* (C-4, C-5, and N-7 of purine ring)



5′-phosphoribosylglycinamide

This reaction is catalyzed by phosphoribosylglycinamide synthetase and requires ATP as the high energy source to drive the reaction.

3. *Introduction of Formyl Group via H_4folate* (Figure 13.2)
This reaction is catalyzed by phosphoribosylglycinamide formyltransferase, which requires N^5,N^{10}-methylidyne H_4folate as the one-carbon donor.

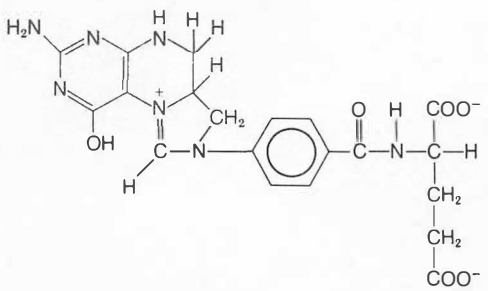
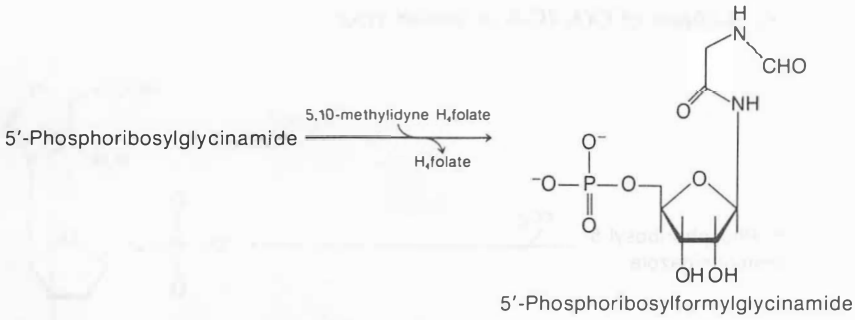
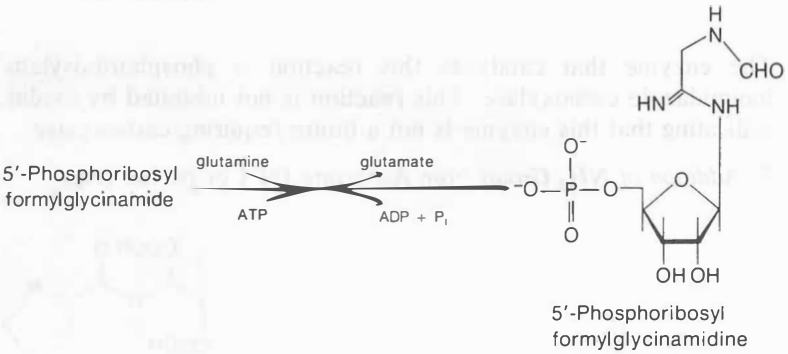


Figure 13.2
Structure of 5,10-methylidyne H_4 folate.

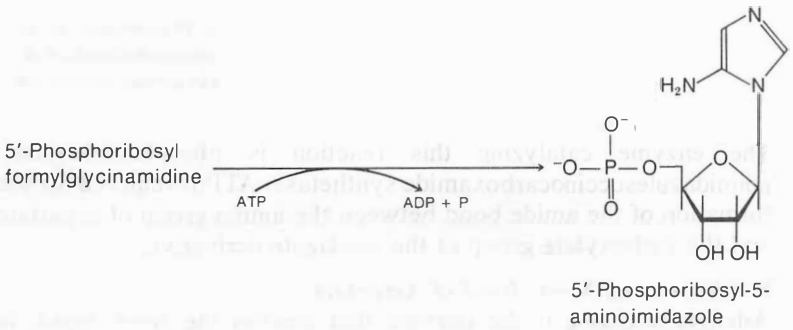


4. *Addition of Nitrogen from Glutamine* (N-3 of purine ring)



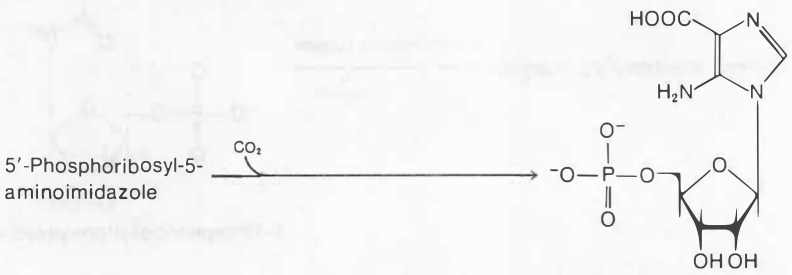
This reaction is catalyzed by phosphoribosylformylglycinamidine synthetase. ATP hydrolysis provides the energy for this reaction.

5. *Ring Closure to Form Imidazole Ring*



The enzyme for this step is phosphoribosylaminoimidazole synthetase, which again requires ATP hydrolysis to close the ring.

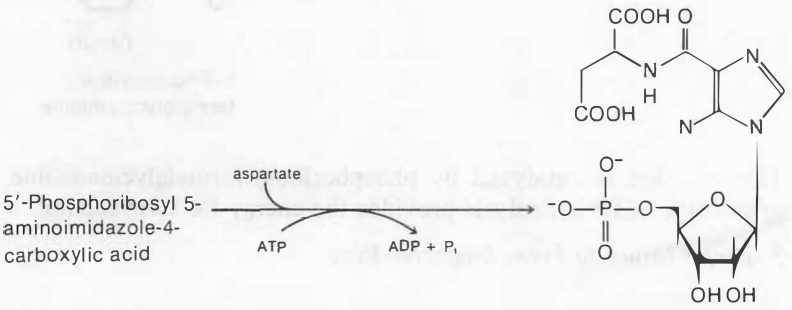
6. *Addition of CO_2* (C-6 of purine ring)



5′-Phosphoribosyl 5-aminoimidazole-4-carboxylic acid

The enzyme that catalyzes this reaction is phosphoribosylaminoimidazole carboxylase. This reaction is not inhibited by avidin, indicating that this enzyme is not a biotin-requiring carboxylase.

7. *Addition of NH_2 Group from Aspartate* (N-1 of purine ring)

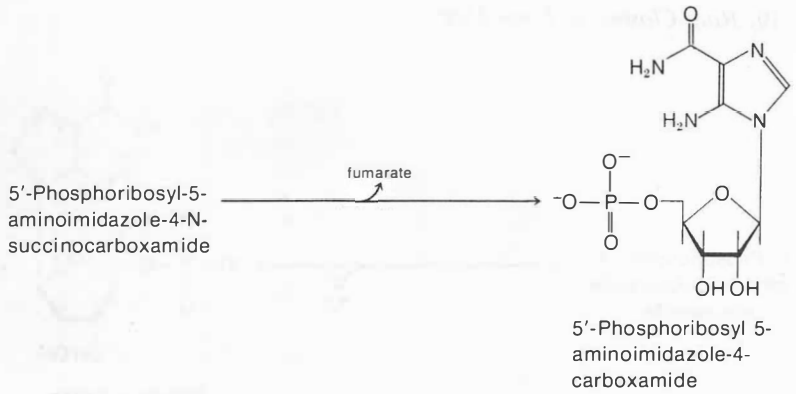


5′-Phosphoribosyl 5-aminoimidazole-4-N-succinocarboxamide

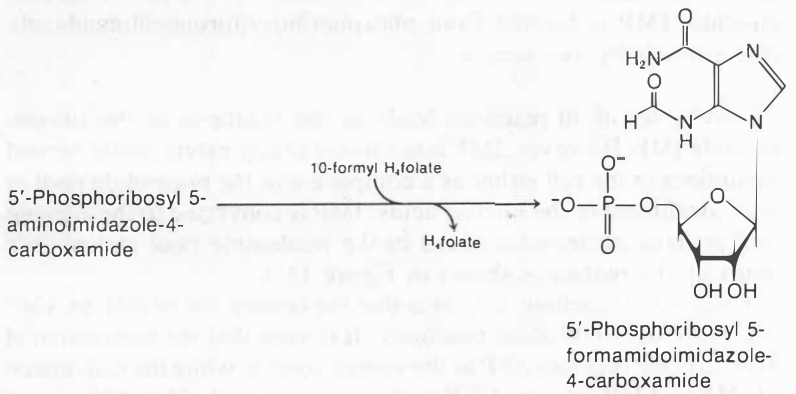
The enzyme catalyzing this reaction is phosphoribosylaminoimidazolesuccinocarboxamide synthetase. ATP is required for the formation of the amide bond between the amino group of aspartate and the carboxylate group of the imidazole derivative.

8. *Cleavage of N—C Bond of Aspartate*

Adenylosuccinase is the enzyme that cleaves the N—C bond, in effect, to transfer the amino group to the imidazole derivative. Adenylosuccinase is the same enzyme that is used in the conversion of adenylosuccinate to AMP and fumarate later in the pathway.



9. *Introduction of Formate via H_4 Folate* (C-2 of purine ring) (Figure 13.3)



The enzyme catalyzing this reaction is phosphoribosylaminoimidazole carboxamide formyltransferase. N^{10}-Formyl H_4folate is the cosubstrate in this reaction and is the one-carbon donor.

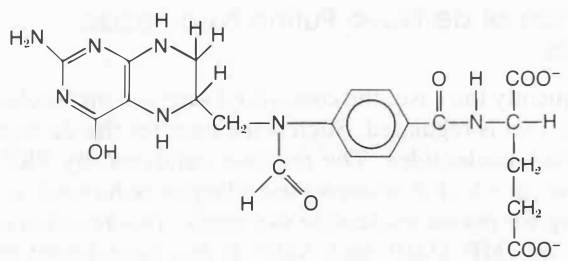
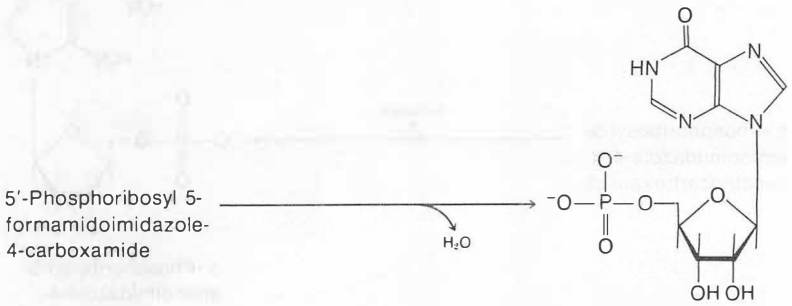


Figure 13.3
Structure of 10-formyl H_4 folate.

10. *Ring Closure to Form IMP*



Inosine 5′-monophosphate (IMP)

In the final reaction of the de novo synthesis of the purine ribonucleotide, IMP is formed from phosphoribosylformamidoimidazole carboxamide by inosinicase.

This series of 10 reactions leads to the synthesis of the ribonucleotide IMP. However, IMP is not found to any extent under normal conditions in the cell either as a component of the nucleotide pool or as a constituent of the nucleic acids. IMP is converted to the adenine and guanine nucleotides found in the nucleotide pool and nucleic acids by the pathways shown in Figure 13.4.

From these reactions it is clear that the conversion of IMP to AMP and GMP does not occur randomly. It is seen that the conversion of IMP to GMP requires ATP as the energy source, while the conversion of IMP to AMP requires GTP as the energy source. Therefore, when there is sufficient ATP in the cell, IMP will be converted to GMP, and conversely, when there is sufficient GTP in the cell IMP will be converted to AMP.

Regulation of de Novo Purine Nucleotide Synthesis

As is frequently the case, the committed step of a metabolic pathway is the step that is regulated. Such is the case for the de novo synthesis of purine nucleotides. The reaction catalyzed by PRPP amidotransferase, in which 5-phosphoribosylamine is formed, is the committed step for purine nucleotide synthesis. This reaction is strongly regulated by IMP, GMP, and AMP. It has been found that PRPP amidotransferase from human placenta exists in two forms with molecular weights of 133,000 and 270,000. The enzyme activity correlated with the smaller form. In the presence of 5′-nucleotides the

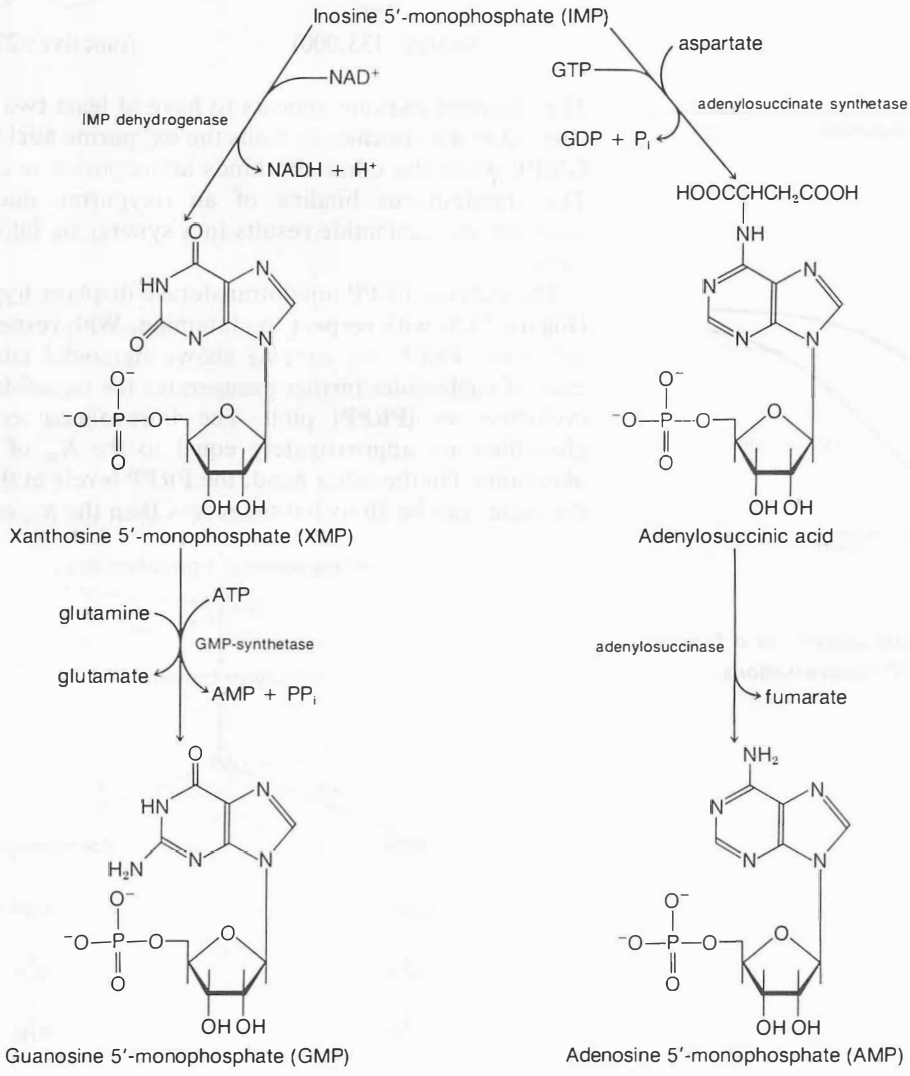


Figure 13.4
Formation of AMP and GMP from IMP branch point.

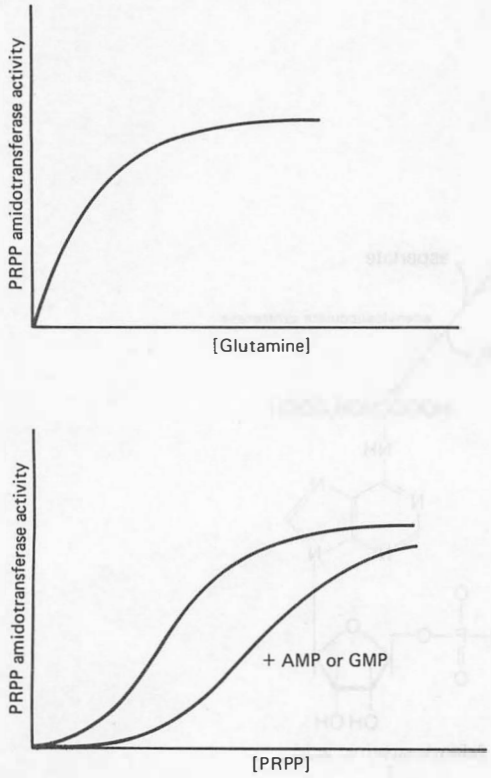
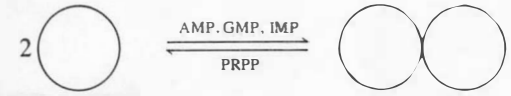


Figure 13.5
PRPP-amidotransferase activity as a function of glutamine or PRPP concentrations.

small active form was converted to the large inactive form, whereas PRPP caused the large form to shift to the active form. This can be viewed in the following manner.



(active; 133,000) (inactive; 270,000)

The placental enzyme appears to have at least two effector binding sites. One site specifically binds the oxypurine nucleotides (IMP and GMP), while the other site binds aminopurine nucleotides (AMP). The simultaneous binding of an oxypurine nucleotide and an aminopurine nucleotide results in a synergistic inhibition of the enzyme.

The enzyme PRPP amidotransferase displays hyperbolic kinetics (Figure 13.5) with respect to glutamine. With respect to the second substrate, PRPP, the enzyme shows sigmoidal kinetics. The presence of nucleotides further exaggerates the sigmoidal nature of the v (velocity) vs [PRPP] plot. The intracellular concentrations of glutamine are approximately equal to the K_m of the enzyme for glutamine. On the other hand, the PRPP levels in the cell, which do fluctuate, can be 10 to 100 times less than the K_m of the enzyme for

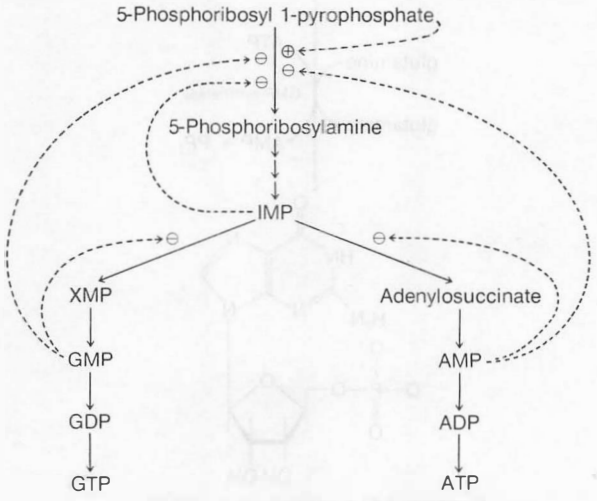


Figure 13.6
Regulation of purine nucleotide synthesis.
The solid arrows indicate enzyme-catalyzed reaction. The dashed arrows indicate the regulated steps (+, activation; −, inhibition).

PRPP. Consequently, the PRPP concentrations in the cell play an important role in controlling de novo purine nucleotide synthesis.

Between the formation of 5-phosphoribosylamine and IMP, there are no known regulated steps. However, there is regulation at the branch point of IMP to GMP and IMP to AMP. The two enzymes which utilize IMP at this branch point, IMP dehydrogenase and adenylosuccinate synthetase, have similar K_m's for IMP. AMP is a competitive inhibitor (with respect to IMP) of adenylosuccinate synthetase, while GMP is a competitive inhibitor of IMP dehydrogenase. Two levels of control are therefore in effect at the IMP branch point. GTP serves as an energy source for the adenylosuccinate synthetase reaction, while AMP is a competitive inhibitor of this step; and ATP serves as the energy source in the conversion of XMP to GMP, while GMP acts as an inhibitor of XMP formation.

The regulation of purine nucleotide synthesis is summarized in Figure 13.6.

Salvage Pathways for the Purine Bases

The efficiency of cellular metabolism under normal conditions is again expressed by the presence of the so-called "salvage pathways" for the purine bases. As seen in the metabolic pathway for the de novo synthesis of the purine nucleotides, a great deal of energy in the form of ATP was required to synthesize the purine nucleotides. In the "salvage" pathways the preformed bases (from exogenous sources or from the turnover of nucleic acids) can be reutilized, resulting in a considerable energy saving for the cell. There are two distinct enzymes involved. The reactions are as follows.

$$\text{Guanine} + \text{PRPP} \rightarrow \text{GMP} + \text{PP}_i$$

$$\text{Hypoxanthine} + \text{PRPP} \rightarrow \text{IMP} + \text{PP}_i$$

The enzyme that catalyzes both these reactions is hypoxanthine-guanine phosphoribosyltransferase (HGPRTase) and requires Mg^{2+}. HGPRTase is regulated by the presence of IMP or GMP. IMP and GMP are competitive inhibitors with respect to PRPP in the HGPRTase reaction. The competitive nature of the inhibition implies that high concentrations of PRPP can overcome the regulation of this metabolic step.

$$\text{Adenine} + \text{PRPP} \rightarrow \text{AMP} + \text{PP}_i$$

The enzyme that catalyzes this reaction is adenine phosphoribosyltransferase (APRTase) and also requires Mg^{2+}. AMP, the product of the reaction catalyzed by APRTase, is an inhibitor of this reaction.

CLIN. CORR. 13.1 LESCH–NYHAN SYNDROME

The Lesch–Nyhan syndrome is characterized clinically by hyperuricemia, excessive uric acid production, and neurological problems, which may include spasticity, mental retardation, and self-mutilation. This disorder is associated with a very severe or complete deficiency of HGPRTase activity. It is an inherited defect that affects only males. In a study of the available patients, it was observed that if the HGPRTase activity was less than 2% of normal, mental retardation was present,

and if the activity was less than 0.2% of normal, the self-mutilation aspect was expressed. This defect leads not only to the overproduction and excretion of uric acid, but also to increased excretion of hypoxanthine.

There appear to be several variant forms of the HGPRTase-deficient patients. In one form there appears to be a complete lack of HGPRTase activity, and yet by titration with antibodies prepared against the purified enzyme from normal erythrocytes, the Lesch–Nyhan patient had an equivalent amount of immunoprecipitable protein. In another form, the enzyme appeared to be unstable. "Young" red cells separated from the "old" red cells by density centrifugation had much higher (although still markedly reduced compared to normal) HGPRTase activity than did the "old" red cells. In still another case, a child was studied who had all the clinical manifestations of the Lesch–Nyhan syndrome, but had normal levels of HGPRTase activity when determined in the laboratory. Further kinetic analysis showed that this variant was a "K_m mutant." When assayed under saturating levels of PRPP, the level of activity was comparable to normal. When assayed at the concentration of PRPP, which would approximate the intracellular concentration of PRPP, the activity was diminished to the range found in Lesch–Nyhan patients.

As discussed previously, the role of HGPRTase is to catalyze reactions in which hypoxanthine and guanine are converted to nucleotides. The hyperuricemia and excessive uric acid production which occur in patients with the Lesch–Nyhan syndrome are easily explained by the lack of HGPRTase activity. The hypoxanthine and guanine are not salvaged leading to increased intracellular pools of PRPP and decreased levels of IMP or GMP. Both these factors promote the de novo synthesis of purine nucleotides.

These reactions are important not only because they conserve energy, but also because they permit cells such as erythrocytes to form nucleotides from the bases. The erythrocyte, for example, does not have PRPP amidotransferase and hence cannot synthesize 5-phosphoribosylamine, the first unique metabolite in the pathway of purine nucleotide synthesis. As a consequence, the red cell must depend on the purine phosphoribosyltransferases to replenish the nucleotide pools.

The importance of these reactions is further demonstrated in the situation in which the HGPRTase activity is markedly depressed. Such a deficiency results in the Lesch–Nyhan syndrome (Clin. Corr. 13.1), which is characterized clinically by hyperuricemia, mental retardation, and self-mutilation.

Interconversion of Purine Nucleotides

Along with the very fine control exhibited by the cell for the de novo synthesis of purine nucleotides, there are enzymes that can be used to balance the levels of guanine and adenine nucleotides. As discussed earlier, IMP can be converted by one pathway to GMP and by a different pathway to AMP. There is no known direct pathway for the conversion of GMP to AMP or AMP to GMP. However, these purine nucleotides can be redistributed to meet the cellular needs through the conversion of GMP and AMP back to IMP. These reactions are carried out by separate enzymes, each under separate controls. The pathways are summarized in Figure 13.7. The reductive deamination of GMP to IMP by GMP reductase is activated by GTP and inhibited by XMP.

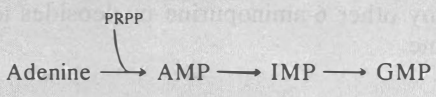
XMP is a competitive inhibitor of human GMP reductase, having a $K_i = \sim 0.2\ \mu M$. Because of this low K_i, the concentration of XMP in the cell could influence the conversion of GMP to IMP. On the other hand, GTP is a nonessential activator of GMP reductase and serves to lower the K_m of the enzyme with respect to GMP and to increase the V_{max}. The activity of AMP deaminase (5′-AMP aminohydrolase), which specifically catalyzes the deamination of AMP to yield IMP, is activated by K^+, and ATP and is inhibited by inorganic phosphate, GDP, and GTP. In the absence of K^+ ions the v vs [AMP] curve is sigmoidal. The presence of K^+ ions is not required for maximum activity but rather acts as a positive allosteric effector to reduce the apparent K_m for AMP.

Degradation of Purine Nucleotides

The purine nucleotides, nucleosides, and bases funnel through a common pathway for the degradation of these biomolecules. The

sis of purine nucleotides without regard for the proper regulation of this pathway.

It is not understood why a severe defect in this salvage pathway leads to the neurological problems. The adenine phosphoribosyltransferase activity in these patients is normal or in fact elevated. With this salvage enzyme, presumably the cellular needs for purine nucleotides could be met via the pathway,



if the cell's de novo pathway were not functioning. The normal tissue distribution of HGPRTase activity perhaps could explain the neurological symptoms. The brain (frontal lobe, basal ganglia, and cerebellum) has 10 to 20 times the enzyme activity found in liver, spleen, or kidney and from 4 to 8 times that found in the erythrocytes. Individuals who have primary gout with excessive uric acid formation and hyperuricemia do not display the neurological problems. It is argued that on this basis the products of purine degradation (hypoxanthine, xanthine, and uric acid) cannot be toxic to the central nervous system (CNS). However, it is possible that these metabolites are toxic to the developing CNS or that the lack of the enzyme leads to an imbalance in the concentrations of the purine nucleotides at critical times during development.

Perhaps the lesion which leads to the neurological problems is not related to the HGPRTase deficiency. It will require further study to understand the relationship between these neurological defects and the decreased levels of HGPRTase.

Treatment of the Lesch–Nyhan patients with allopurinol will decrease the amount of uric acid formed, relieving some of the problems caused by sodium urate deposits. However, since the Lesch–Nyhan patient has a marked reduction in HGPRTase activity, hypoxanthine and guanine are not salvaged, PRPP is not consumed, and consequently the de novo synthesis of purine nucleotides is not shut down. There is no known treatment for the neurological problems. These patients usually die from kidney failure, resulting from the high sodium urate deposits.

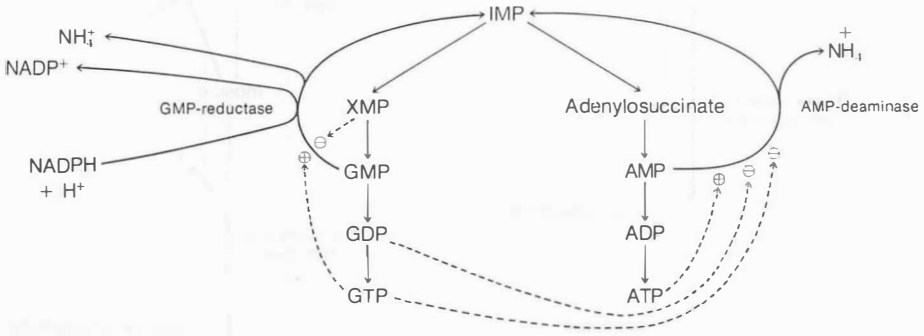


Figure 13.7
Interconversion of purine nucleotides.
The solid arrows represent enzyme-catalyzed reactions. The dashed lines represent steps of regulation (⊕, activation; ⊖, inhibition).

end product of purine degradation in man is uric acid. The catabolic pathways are as shown in Figure 13.8. The enzymes involved in the degradation of the nucleic acids and the nucleotides and nucleosides vary in specificity. The nucleases show specificity toward either RNA or DNA and also toward the bases and position of cleavage of the 3′,5′-phosphodiester bonds. The nucleotidases range from those with relatively high specificity, such as 5′-AMP nucleotidase, to those with broad specificity, such as the acid and alkaline phosphatases, which will hydrolyze any of the 3′- or 5′-nucleotides. AMP deaminase is specific for AMP. Adenosine deaminase is much less specific, since not only adenosine, but also 2′-deoxyadenosine and many other 6-aminopurine nucleosides are deaminated by this enzyme.

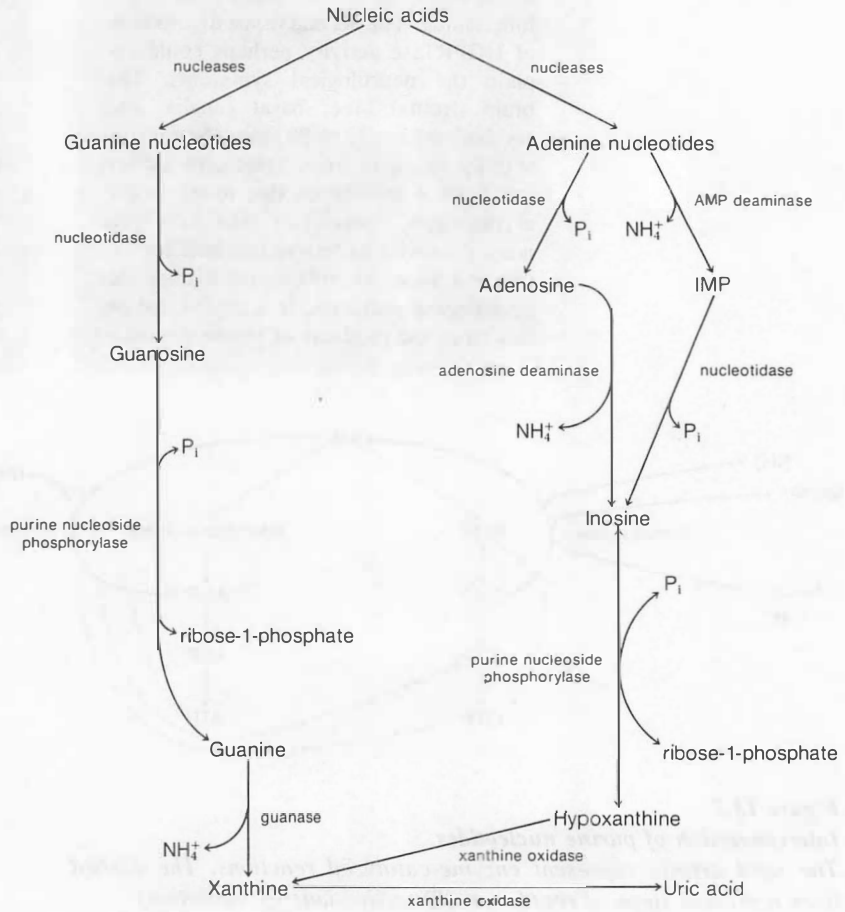


Figure 13.8
Degradation of purine nucleotides.

A special comment should be made about purine nucleoside phosphorylase. As indicated, the reaction is readily reversible.

$$\text{Inosine} + P_i \rightleftharpoons \text{hypoxanthine} + \text{ribose 1-P}$$

or

$$\text{Guanosine} + P_i \rightleftharpoons \text{guanine} + \text{ribose 1-P}$$

or

$$\text{Xanthosine} + P_i \rightleftharpoons \text{xanthine} + \text{ribose 1-P}$$

Deoxyinosine and deoxyguanosine are also excellent substrates for purine nucleoside phosphorylase. While the equilibrium for the reactions catalyzed by purine nucleoside phosphorylase favors nucleoside synthesis, it would appear that in the cell the concentrations of the free purine and ribose 1-phosphate are too low to support nucleoside synthesis under most conditions. The main function of purine nucleoside phosphorylase in the cells seems to be its role in purine nucleoside degradation. This is supported by the conditions observed in the cases where a deficiency of purine nucleoside phosphorylase has been detected. Under these conditions there is a large buildup of the substrates (inosine, guanosine, deoxyinosine and deoxyguanosine) for purine nucleoside phosphorylase with a corresponding decrease in uric acid formation.

There is some evidence that a separate enzyme (with low activity or present in low concentrations) is present for the phosphorolysis of adenosine.

These enzymes may therefore be regarded as part of the degradative pathway. On the other hand, since the reaction is readily reversible, it may serve as part of the salvage pathway under certain metabolic conditions.

Deficiencies in two of the enzymes of this degradative pathway (adenosine deaminase and purine nucleoside phosphorylase) have been observed in man in two diseased states. Adenosine deaminase deficiency has been associated with a severe combined immunodeficiency, and purine nucleoside phosphorylase deficiency is associated with a defective T-cell immunity and a normal B-cell immunity (Clin. Corr. 13.2).

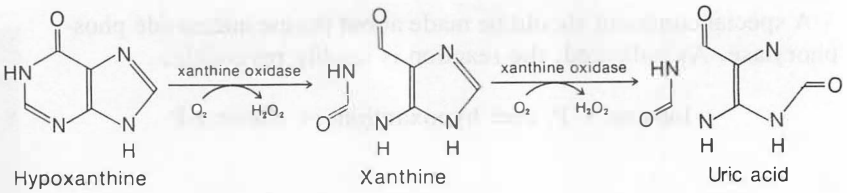
Formation of Uric Acid

Xanthine oxidase is an enzyme which contains FAD, Fe(III), and Mo(VI). In the reaction, molecular oxygen is a substrate with H_2O_2 being generated as a product. Uric acid is the end product of purine

CLIN. CORR. 13.2 IMMUNODEFICIENCY DISEASES ASSOCIATED WITH DEFECTS IN PURINE NUCLEOTIDE METABOLISM

Recently two different immunodeficiency diseases have been recognized that are associated with deficiencies in the enzymes adenosine deaminase and purine nucleoside phosphorylase. Referring to Figure 13.8, it is seen that these two enzymes are involved in the degradation of purine nucleosides.

The deficiency in adenosine deaminase is associated with a severe combined immunodeficiency involving T-cell and usually B-cell dysfunction. Adenosine deaminase deficiency is not associated with the overproduction of purine nucleotides. The mechanism by which the lack of adenosine deaminase interferes with immune function is not completely understood. However, it has been shown that in a patient with adenosine deaminase deficiency there is an extremely large buildup of deoxyadenosine triphosphate in the erythrocytes examined. In fact, the dATP concentration exceeded the ATP concentration in these cells. It is therefore thought that the failure of the cells to metabolize deoxyadenosine to deoxyinosine for further conversion to hypoxanthine leads to the increased levels of dATP. dATP is known to be a very effective inhibitor of ribonucleotide reductase and consequently of DNA synthesis (cell replication). It is thought that this is the site that leads to the deficiencies in the immune system. Other suggestions have included the proposals that the elevated adenosine is toxic to the cells by virtue of



its ability to increase the intracellular concentrations of cAMP or due to the inhibition of *S*-adenosine homocysteine hydrolase, leading to increased intracellular levels of *S*-adenosyl-L-homocysteine.

The deficiency in purine nucleoside phosphorylase is associated with an impairment of T-cell function with no apparent effects on B-cell function. There is no overproduction of purine nucleotides associated with this deficiency. However, there is a marked *decrease* in uric acid formation with the corresponding increased levels of the purine nucleoside phosphorylase substrates, guanosine, deoxyguanosine, inosine, and deoxyinosine. When these various nucleosides were incubated with normal T lymphocytes in culture, deoxyguanosine was found to be the most toxic. In addition, it was found that dGTP was the major nucleotide that accumulated in the red cells from patients with purine nucleoside phosphorylase deficiency. It is suggested that dGTP, which acts as an inhibitor of CDP reductase, is the actual agent which is toxic to the development of normal T cells.

In both of these enzyme deficiencies, it is not entirely clear how these defects lead to the immune problems. However, it is clear that defects in enzymes that have been casually considered in metabolic pathways in the past, reveal their importance to normal metabolism when they are absent or severely decreased.

nucleotide catabolism and is excreted in the urine. There are clinical disorders in which the serum level of uric acid is markedly elevated (hyperuricemia) and which can lead to the deposit of sodium urate crystals. This condition is known as gout (Clin. Corr. 13.3).

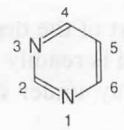
Allopurinol is an inhibitor of xanthine oxidase; this is discussed in detail as a clinical drug in Clin. Corr. 13.3.

13.5 METABOLISM OF PYRIMIDINE NUCLEOTIDES

De Novo Synthesis

The pyrimidine ring is synthesized de novo in mammalian cells utilizing amino acids as carbon and nitrogen donors and CO_2 as a carbon donor.

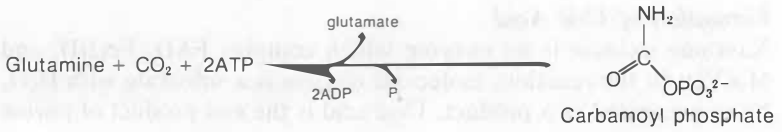
The numbering system for the pyrimidine ring is shown below along with the sources for the various carbon and nitrogen atoms indicated.



N-1, C-4, C-5, and C-6 from aspartate
C-2 from CO_2
N-3 from amide N of glutamine

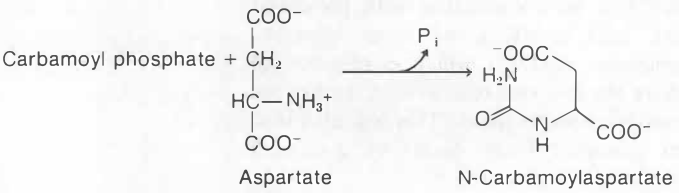
As in the case of the de novo synthesis of purine nucleotides, amino acids also play an important role in the de novo synthesis of pyrimidine nucleotides. The reactions leading to the synthesis of pyrimidine nucleotides in mammalian cells are as follows.

1. *Formation of Carbamoyl Phosphate* (C-2 and N-3 of pyrimidine ring)



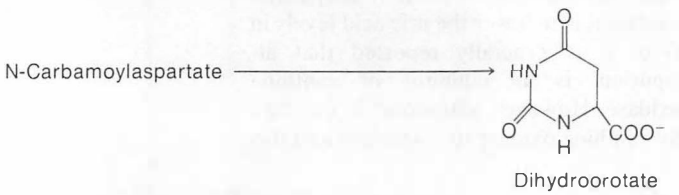
The enzyme that catalyzes this reaction is called carbamoyl phosphate synthetase II. It is distinct from carbamoyl phosphate synthetase I, a mitochondrial enzyme involved in the urea cycle.

2. *Addition of Aspartate* (N-1, C-4, C-5, and C-6 of pyrimidine ring)



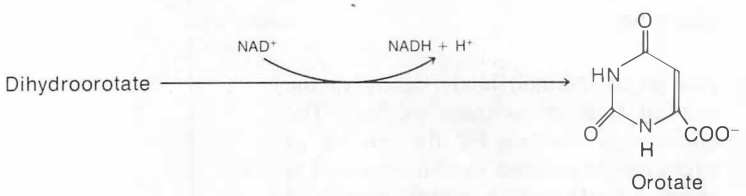
This reaction is catalyzed by the enzyme aspartate carbamoyltransferase. The mammalian enzyme is not allosterically regulated, although this could be considered to be the committed step for pyrimidine synthesis.

3. *Ring Closure to Form Pyrimidine Ring*



This reaction is catalyzed by the enzyme dihydroorotase.

4. *Oxidation of Dihydroorotate*



Dihydroorotate dehydrogenase, a flavoprotein, catalyzes the reaction in which orotic acid is formed.

5. *Addition of the Ribose 5-Phosphate* (formation of N-riboside bond) The enzyme catalyzing the reaction in which the first pyrimidine nucleotide is formed is orotate phosphoribosyltransferase. PRPP is the ribose-5-phosphate donor.

CLIN. CORR. 13.3 GOUT

Primary gout is characterized by excessive uric acid due to a variety of metabolic abnormalities that lead to the overproduction of purine nucleotides via the de novo pathway. With the overproduction of uric acid, the levels of uric acid in the serum are elevated, and there are deposits of sodium urate crystals in the joints of extremities. The metabolic bases for the increased production of purine nucleotides which in turn is manifested by increased uric acid levels have been identified in several situations. Many, if not all, of the clinical symptoms associated with the overproduction of uric acid arise because uric acid is not very soluble. Formation of sodium urate crystals leads not only to the joint problems but also to renal disease. Hyperuricemia resulting from the overproduction of uric acid via the de novo pathway as opposed to hyperuricemia resulting from renal damage or increased cell death (e.g., radiation therapy) can be relatively easily distinguished. The feeding of ^{15}N-glycine to an "overproducer" will result in a marked ^{15}N enrichment of the N-7 of uric acid isolated from the urine or serum of these patients, whereas there would be little ^{15}N enrichment in the uric acid from individuals with renal problems or with increased nucleic acid degradation.

Various studies directed at determining the molecular basis in primary gout for the overproduction of purine nucleotides, and consequently of uric acid, have uncovered a diverse group of metabolic defects. While the primary metabolic defects that have been determined may be seemingly unrelated to the de novo synthesis of purine nucleotides, a common feature of these defects evolves.

The defects described in human beings include the following:

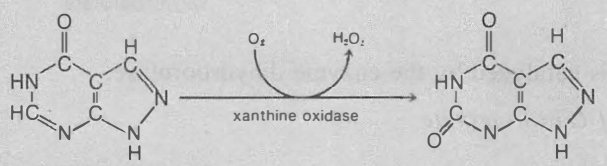
1. *PRPP-Synthetase:* Mutant forms of PRPP-synthetase have been detected, which are not subject to allosteric regulation by inorganic phosphate or feedback inhibition by GDP and ADP. Under these conditions, the intracellular concentration of PRPP is elevated, leading to increased formation of 5-phosphoribosylamine.
2. *Partial HGPRTase Deficiency:* A characteristic of this deficiency is the overproduction of purine nucleotides. The basis for this appears to be twofold. First, the lack of HGPRTase activity decreases the amount of hypoxanthine or guanine that can be "salvaged." Consequently, the level of PRPP is increased because PRPP is not consumed via the salvage enzyme. The increased PRPP levels lead to increased PRPP amidotransferase activity. Second, the lack of salvage of hypoxanthine or guanine leads to decreased levels of IMP and GMP, which in turn act as feedback regulators of the PRPP amidotransferase step.

In both of these conditions, the common feature of the defect that leads to the overproduction of purine nucleotides is that the intracellular concentration of PRPP is elevated. These defined defects fully support the conclusion that the PRPP amidotransferase step is the rate-controlling step in purine nucleotide synthesis.

Further support comes from clinical cases of secondary gout (in consequence of another metabolic defect). A deficiency in glucose 6-phosphatase (glycogen storage disease, type I; von Gierke's) leads to increased purine nucleotide synthesis de novo. The lack of conversion of glucose 6-phosphate to glucose leads to increased hexose monophosphate shunt activity. The increased utilization of glucose 6-phosphate via the shunt results in increased ribose 5-phosphate levels and consequently increased PRPP levels. An elevation of glutathione reductase activity has also been correlated with increased uric acid levels. Glutathione reductase generates $NADP^+$, which is required to drive the first two reactions of the hexose monophosphate shunt. This will also lead to increased PRPP levels as discussed above.

These latter two examples show quite clearly that a defect in one pathway can cause major problems in a metabolic pathway that is seemingly unrelated and points to the critical nature of the interrelationships among various pathways.

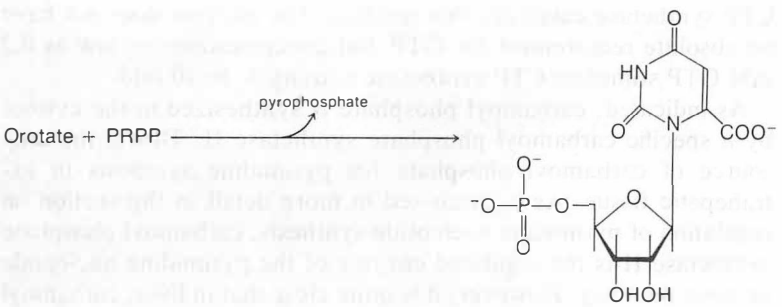
Probably the major treatment for primary gout involves the use of the drug allopurinol. The overall effect of allopurinol treatment is to lower the uric acid levels in vivo. It is generally reported that allopurinol is an inhibitor of xanthine oxidase. However, allopurinol is oxidized by xanthine oxidase to oxipurinol and this



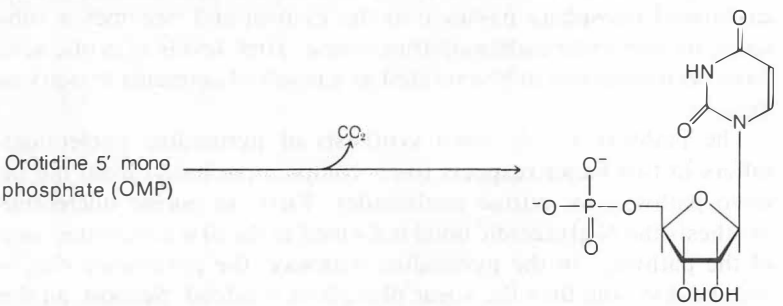
Allopurinol Oxipurinol

product (oxipurinol) binds tightly to the reduced form of xanthine oxidase. The dissociation constant for the binding of oxipurinol to reduced xanthine oxidase is about 0.5 nM, which makes oxipurinol a very effective inhibitor of xanthine oxidase.

The inhibition of xanthine oxidase by oxipurinol decreases the formation of uric acid, while increasing the levels of hypoxanthine and xanthine excreted. This benefits the patient, since the amount of

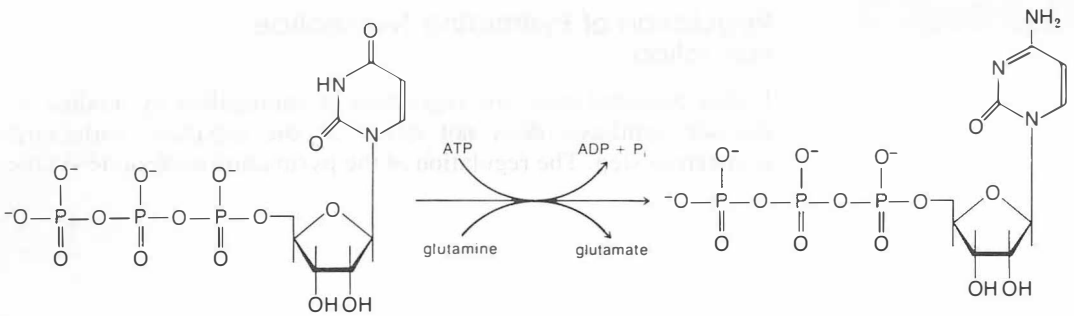


6. *Decarboxylation of OMP*



OMP-decarboxylase catalyzes this reaction. The absence of both or either of these two last enzymes leads to a condition termed orotic aciduria (Clin. Corr. 13.4)

By these reactions the pyrimidine nucleotide UMP is synthesized. The formation of cytidine nucleotides proceeds from the uridine nucleotide but at the triphosphate level rather than at the monophosphate level. This reaction for the synthesis of this pyrimidine is as follows:



the purine degradative products will be distributed among three compounds instead of just uric acid. Hypoxanthine and xanthine are more soluble than uric acid, so that the total amount (hypoxanthine, xanthine, and uric acid) that will be soluble is increased by allopurinol treatment.

In "overproducers" who do not have a deficiency in HGPRTase, allopurinol treatment not only lowers the formation of uric acid with an increase in the excretion of hypoxanthine and xanthine, but also decreases the overall production of purine nucleotides via the de novo pathway. The metabolic basis for this appears to be due to the increased salvage of hypoxanthine and xanthine, which requires the consumption of PRPP and the subsequent formation of IMP and XMP, which can block de novo synthesis at the PRPP amidotransferase step.

CLIN. CORR. 13.4 OROTIC ACIDURIA

In a clinical condition called hereditary orotic aciduria, characterized by retarded growth and severe anemia, high levels of orotic acid are excreted. The biochemical basis for this increased production of orotic acid is the absence of either or both of the enzymes orotate phosphoribosyltransferase and OMP-decarboxylase. While this is a relatively rare disease, the understanding of the metabolic basis for it has led not only to a successful treatment, but also has confirmed the site of regulation of pyrimidine nucleotide synthesis.

When these patients are fed either cytidine or uridine there is not only a marked improvement in the hematologic manifestation but also a decrease in orotic acid formation. The biochemical basis for this can be explained as follows. Uridine or cytidine after conversion to the nucleotide by the cell bypasses the block at orotate phosphoribosyltransferase/OMP-decarboxylase to provide the rapidly growing cells, such as the erythropoietic cells, with the pyrimidine nucleotides required for RNA and DNA synthesis. In addition, the intracellular formation of UTP from these nucleosides acts as a feedback inhibitor of carbamoyl phosphate synthetase II to shut down the synthesis of orotic acid.

CTP synthetase catalyzes this reaction. The enzyme does not have an absolute requirement for GTP, but concentrations as low as 0.2 mM GTP stimulate CTP synthetase activity 5- to 10-fold.

As indicated, carbamoyl phosphate is synthesized in the cytosol by a specific carbamoyl phosphate synthetase II. This is the only source of carbamoyl phosphate for pyrimidine synthesis in extrahepatic tissue. As is discussed in more detail in the section on regulation of pyrimidine nucleotide synthesis, carbamoyl phosphate synthetase II is the regulated enzyme of the pyrimidine nucleotide de novo pathway. However, it is quite clear that in liver, carbamoyl phosphate synthetase I can provide cytosolic carbamoyl phosphate. Under stressed physiological conditions in which there is excessive ammonia, the liver, through carbamoyl phosphate synthetase I (a component of the urea cycle) can utilize this pathway to detoxify the NH_3 by forming carbamoyl phosphate in the mitochondria. This carbamoyl phosphate passes into the cytosol and becomes a substrate for aspartate carbamoyltransferase. High levels of orotic acid have been observed to be excreted as a result of ammonia toxicity in man.

The pathway for de novo synthesis of pyrimidine nucleotides differs in two major respects (on a comparative basis) from the de novo pathway for purine nucleotides. First, in purine nucleotide synthesis the N-glycosidic bond is formed in the first committed step of the pathway. In the pyrimidine pathway, the pyrimidine ring is formed first, and then the sugar phosphate is added. Second, all the enzymes of the purine nucleotide pathway are in the cytosol. For pyrimidine nucleotide synthesis one of the enzymes, dihydroorotate dehydrogenase, is found in the mitochondria. The other enzymes are cytocolic and are found as complexes. Carbamoyl phosphate synthetase II, aspartyl transcarbamoylase, and dihydroorotase make up one complex, while orotate phosphoribosyltransferase and OMP-decarboxylase are found in a second complex. As a result of the channeling of intermediates in these enzyme complexes, none of the intermediates are normally found in the acid-soluble fractions prepared from cells.

Regulation of Pyrimidine Nucleotide Formation

Unlike bacterial cells, the regulation of mammalian pyrimidine nucleotide synthesis does not occur at the aspartate carbamoyltransferase step. The regulation of the pyrimidine nucleotide synthe-

sis in mammalian cells occurs at the level of carbamoyl phosphate synthetase II (the cytosolic enzyme), which is inhibited by UTP.

The next level of regulation of pyrimidine nucleotide synthesis is at the level of OMP-decarboxylase. UMP and, to a lesser extent, CMP are inhibitors of OMP-decarboxylase but not orotate phosphoribosyltransferase. However, since these two enzymes comprise the components of a complex, orotate phosphoribosyltransferase cannot continue to function, since there is no place to transfer the product. Under the conditions in which OMP-decarboxylase is inhibited, orotate, *not* OMP accumulates.

In a clinical condition called orotic aciduria excessive amounts of orotic acid are produced. This is caused by deficiencies in either orotate phosphoribosyltransferase or OMP decarboxylase or both enzymes (Clin. Corr. 13.4).

Another possible regulatory site is the CTP synthetase step. This enzyme shows a hyperbolic curve for a plot of velocity (v) vs [UTP]. However, in the presence of CTP, the plot of v vs [UTP] curve becomes sigmoidal. In this way the activity of CTP synthetase is depressed, preventing all of the UTP from being converted to CTP.

Salvage Pathways for Pyrimidines

Pyrimidines can be "salvaged" by conversion to the nucleotide level by reactions involving pyrimidine phosphoribosyltransferase.

The general reaction is

$$\text{Pyrimidine} + \text{PRPP} \longrightarrow \text{pyrimidine nucleoside monophosphate} + \text{pyrophosphate}$$

The enzyme from human erythrocytes has been purified and can utilize orotate, uracil, and thymine as substrates. Cytosine is not a substrate.

It should be mentioned that as uracil becomes available to the cell, competing reactions can occur. Uracil can be degraded to β-alanine or can be salvaged. Normal liver, when presented with uracil, will readily degrade it, whereas regenerating liver would convert the uracil to UMP. This is the result of the availability of PRPP, the enzyme levels, and in general the metabolic state of the animal.

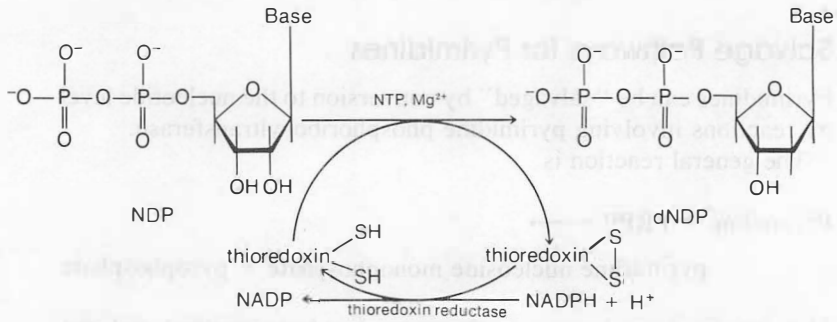
The pyrimidine nucleoside kinases can also be thought of as "salvage" enzymes. The net effect of the kinase reaction is to divert the pyrimidine nucleoside from the degradative pathway to the pyrimidine nucleotide level for cellular utilization.

13.6 DEOXYRIBONUCLEOTIDE FORMATION

As indicated earlier in the chapter, the concentrations of deoxyribonucleotides are extremely low in the "resting" cell. Only at the time of DNA replication (S phase) does the deoxyribonucleotide pool increase to support the required DNA synthesis.

Reduction of Ribonucleotides

Deoxyribonucleotides are formed by the direct reduction of the 2′ position of the corresponding ribonucleotides. The reaction is strongly regulated not only by allosteric effectors (both activators and inhibitors) but also by drastic changes in the level of the enzyme catalyzing the formation of deoxyribonucleotides. This reaction occurs at the level of the nucleoside diphosphate. The general reaction can be summarized as follows:



The enzyme catalyzing the formation of 2′-deoxyribonucleotides is nucleoside diphosphate reductase (ribonucleotide reductase). The reduction of a specific NDP requires a specific NTP as a positive effector of the enzyme. This reaction is also subject to regulation by other NTPs, which can serve as negative effectors. The specificity of the effectors for the various substrates are summarized in Table 13.4.

The mammalian ribonucleotide reductase enzyme consists of two nonidentical subunits neither of which alone has enzymatic activity. One of the subunits contains the effector binding site or sites, while the other subunit contains the nonheme iron. For the mammalian system it has not been completely resolved whether there is one enzyme for all four substrates, or whether there are separate enzymes or binding sites for each of the substrates. dATP is a potent

Table 13.4 Effectors of Ribonucleotide Reductase Activity

Substrate	*Major Positive Effector*	*Major Negative Effectors*
CDP	ATP	dATP, dGTP, dTTP
UDP	ATP	dATP, dGTP
ADP	dGTP	dATP, ATP
GDP	dTTP	dATP

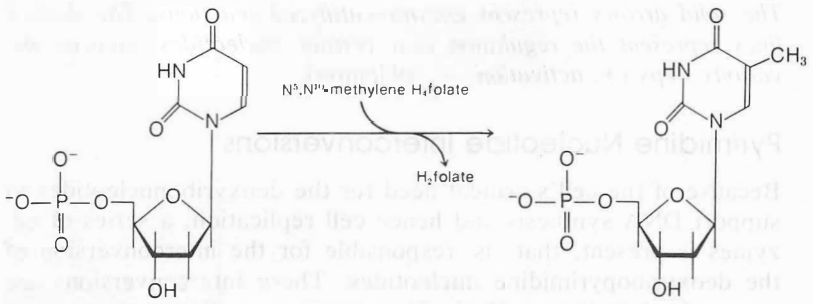
inhibitor of the reduction of all four nucleoside diphosphate substrates. This fact provides the biochemical basis for the toxicity of deoxyadenosine for a variety of mammalian cells.

Thioredoxin is a small molecular weight protein (12,000 daltons), which is oxidized during the reduction of the 2′-hydroxyl group of the ribose moiety. To complete the catalytic cycle, reduced thioredoxin is regenerated by thioredoxin reductase (a flavoprotein) and NADPH.

The importance of this enzyme for DNA replication cannot be overemphasized. Ribonucleotide reductase is uniquely responsible for catalyzing the reactions in which the deoxynucleoside triphosphates are generated for DNA replication. It appears that in controlling the deoxyribonucleotide levels in the cell, at least two approaches are utilized by the cell. These are (1) the actual concentration of reductase in the cells and (2) the very strict allosteric regulation of enzyme activity by the nucleoside triphosphates.

Synthesis of Deoxythymidine Nucleotide

Deoxythymidylate is formed in a unique reaction. The enzyme thymidylate synthetase catalyzes this reaction in which a C_1 unit is not only transferred but also reduced to a methyl group. In the



Deoxyuridine 5′-monophosphate (dUMP)

Thymidine 5′-monophosphate (TMP)

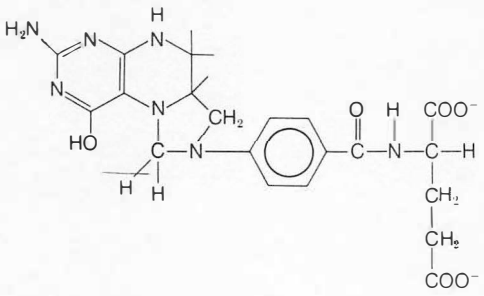


Figure 13.9
Structure of 5,10-methylene H_4folate.

process H_4folate acts both as the C_1 transfer agent and as the reducing agent. As a result dihydrofolate is generated (See Figure 13.9.)

The dUMP for this reaction can arise from two different pathways. In one reaction dCMP is deaminated directly to dUMP by dCMP deaminase, whereas in the other pathway UDP is reduced to dUDP which is then converted to dUMP (Figure 13.10). From labeling studies it appears that for most cells the major source of dUMP is the step in which dCMP is deaminated to dUMP. dCMP deaminase activity in the mammalian cell far exceeds the level of UDP reductase activity.

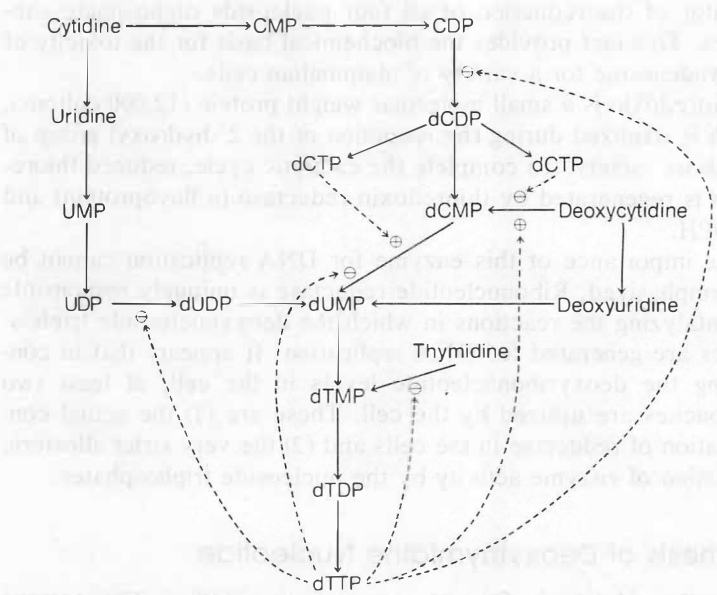


Figure 13.10
Pyrimidine interconversions.
The solid arrows represent enzyme-catalyzed reactions. The dashed lines represent the regulation that certain nucleotides have on the various steps (+, activation; −, inhibition).

Pyrimidine Nucleotide Interconversions

Because of the cell's critical need for the deoxyribonucleotides to support DNA synthesis and hence cell replication, a series of enzymes is present, that is responsible for the interconversion of the deoxyribopyrimidine nucleotides. These interconversions are summarized in Figure 13.10. The enzymes catalyzing these interconversions are strongly regulated. The control of dCDP and dUDP

formation through the ribonucleotide reductase reaction has already been discussed. dCMP deaminase is activated by dCTP and inhibited by dTTP. The inhibition by dTTP can be overcome by increasing concentrations of dCTP. A v vs [dCMP] curve is sigmoidal for dCMP deaminase. The presence of dCTP shifts the curve to a hyperbolic activity curve. dTTP is also an inhibitor of thymidine kinase, an enzyme which is important in "salvaging" thymidine and deoxyuridine for the cell. It is seen that dTTP serves as a negative effector of several conversions, and this is the basis for the toxicity of high concentrations of thymidine for mammalian cells.

Degradation of Pyrimidine Nucleotides

The turnover of nucleic acids results in the release of pyrimidine nucleotides. These nucleotides are also in a steady state, and there is constant synthesis and degradation. The degradation of pyrimidine nucleotides follows the pathways shown in Figure 13.11. In these degradative pathways, the nucleotides are first converted to the nucleoside and then to the free base uracil or thymine. The conversion of the pyrimidine nucleotides to nucleosides is catalyzed by various

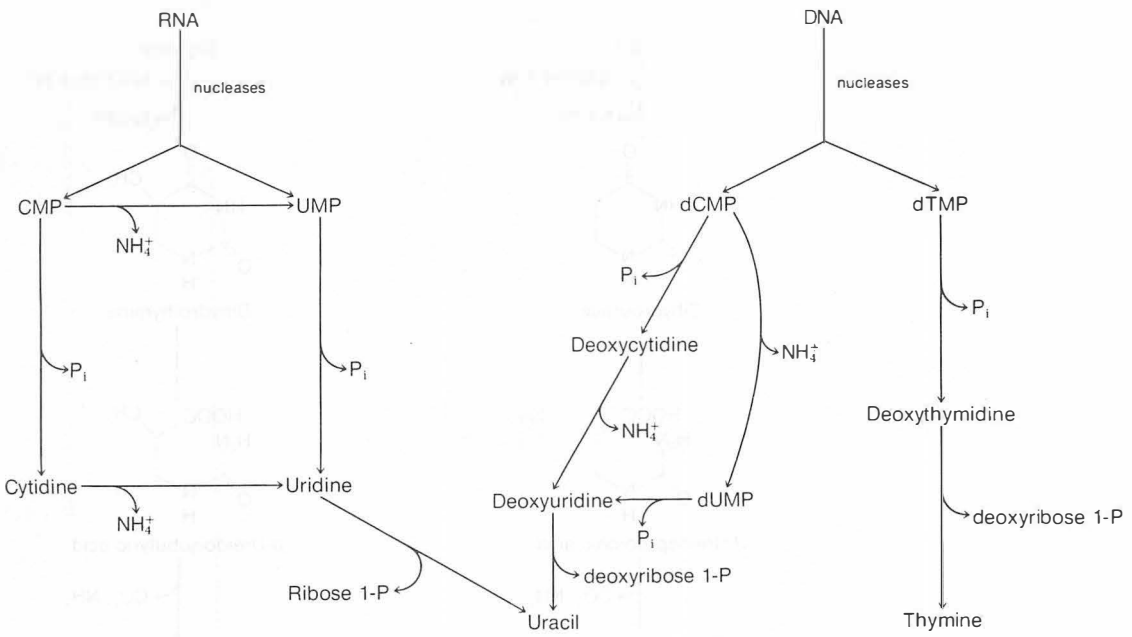


Figure 13.11
Pathways for the degradation of pyrimidine nucleotides.

nonspecific phosphatases. Deoxycytidylate deaminase has preference for dCMP but can utilize CMP as substrate. Cytidine and deoxycytidine are deaminated to uridine and deoxyuridine, respectively, by the nucleoside deaminase. Uridine phosphorylase catalyzes the phosphorolysis not only of uridine, but also deoxyuridine and deoxythymidine.

It is important to note that mammalian cells have a very specific dUTPase in high concentration, which catalyzes the reaction:

$$\text{dUTP} \rightarrow \text{dUMP} + \text{pyrophosphate}$$

It is critical for normal DNA replication that dUTP not be present in the nucleotide pool. As is shown in Chapter 17, dUTP can very effectively replace dTTP as a substrate in the DNA polymerase reaction. dUTPase, therefore, serves to prevent misincorporation of dUMP into the DNA, which would have other major consequences.

Uracil and thymine are then further degraded by analogous reactions, although the final products are different as shown in Figure 13.12. Uracil is degraded further to β-alanine, NH_4^+ and CO_2.

None of these products is unique to uracil degradation, and consequently the turnover of cytosine or uracil nucleotides cannot be

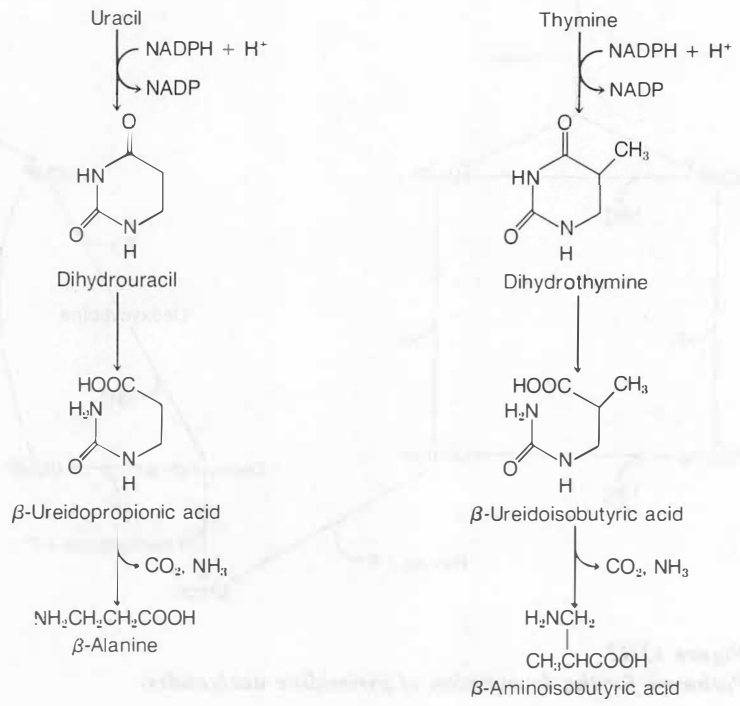


Figure 13.12
Degradation of uracil and thymine.

estimated from the end products of this pathway. Thymine degradation proceeds to β-aminoisobutyric acid, NH_4^+ and CO_2. β-Aminoisobutyric acid is excreted in the urine of man and originates from the degradation of thymine. Increased levels of β-aminoisobutyric acid are excreted after the administration of diets rich in DNA or in cancer patients undergoing chemotherapy or radiation therapy in which large numbers of cells are being killed. It is possible therefore to estimate the turnover of DNA or thymidine nucleotides by the measure of β-aminoisobutyric acid production.

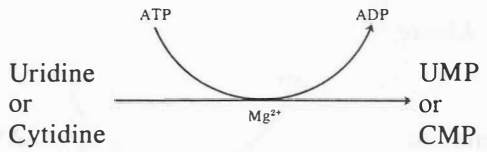
13.7 NUCLEOSIDE AND NUCLEOTIDE KINASES

As shown in the pathways for the de novo synthesis of purine and pyrimidine nucleotides, the nucleotide is synthesized as the monophosphate. Most, if not all, reactions in which the nucleotides function require that these nucleotides be at the di- or triphosphate level.

There are specific kinases to "salvage" nucleosides to nucleotides and to convert the nucleoside monophosphates to the di- and triphosphates.

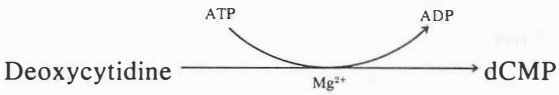
Examples of these are as follows.

1. *Uridine/Cytidine Kinase*



The enzyme is specific for uridine or cytidine as substrates. UTP and CTP are inhibitors of these reactions.

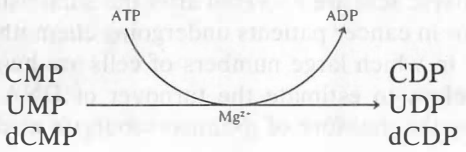
2. *Deoxycytidine Kinase*



Deoxycytidine is the preferred substrate for this kinase. Cytidine, uridine, and thymidine are not substrates. However, deoxyadenosine and deoxyguanosine, although poor substrates, are

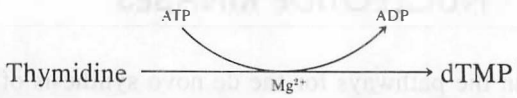
phosphorylated by this enzyme. dCTP is a potent inhibitor of this reaction. dTTP will reverse the inhibition caused by dCTP.

3. *Pyrimidine Nucleoside Monophosphate Kinase*



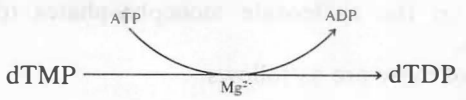
This enzyme shows specificity for the substrates CMP, UMP, and dCMP. dUMP is not a substrate.

4. *Thymidine Kinase*



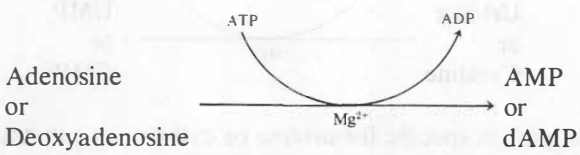
This kinase is specific for thymidine and deoxyuridine as substrates. This enzyme is elevated in rapidly growing tissues.

5. *Thymidylate Kinase*



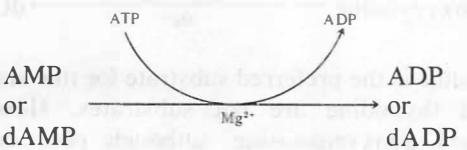
This enzyme is specific for dTMP.

6. *Adenosine Kinase*



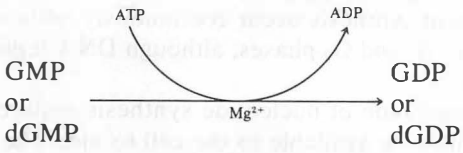
This enzyme is specific for adenosine or deoxyadenosine. Inosine is not a substrate for this enzyme.

7. *AMP Kinase*



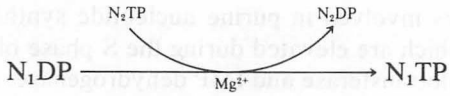
AMP kinase shows specificity for AMP. Although dAMP can be utilized as substrate, AMP is phosphorylated at a rate 10 times higher than that of dAMP.

8. *GMP Kinase*



GMP kinase is distinct from AMP kinase showing specificity for GMP and dGMP.

9. *Nucleoside Diphosphokinase*



Mammalian cells contain an enzyme, nucleoside diphosphokinase, which is not specific for either the phosphate donor or phosphate acceptor in terms of either the purine or pyrimidine base or the sugar moiety. Since in most cells the concentration of ATP is the highest of the triphosphates and most easily regenerated via glycolysis or oxidative phosphorylation, ATP is probably the major phosphate donor for these reactions in intact cells.

13.8 NUCLEOTIDE METABOLIZING ENZYMES AS A FUNCTION OF THE CELL CYCLE AND RATE OF CELL DIVISION

For normal cell division to occur, essentially all components of the cell must double. The events which lead from the formation of a daughter cell through mitosis, to the completion of the processes required for its own division into two daughter cells is described by the term cell cycle. The periods of the cell cycle have been termed mitosis (M), gap 1 (G_1), synthesis (S), and gap 2 (G_2). The total period of the cell cycle will vary with the particular cell type. In many mammalian cells the periods of M, S, and G_2 are relatively

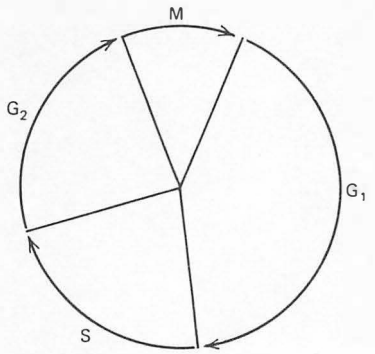


Figure 13.13
Diagrammatic representation of the cell cycle. *For a mammalian cell with a doubling time of 24 h, G_1 would consume ~12 h; S, 7 h; G_2, 4 h; and M, 1 h.*

constant, while G_1 varies widely causing cells to have long or short doubling times. The cell cycle is represented in Figure 13.13. In preparation for DNA replication during the S phase of the cell cycle, there is considerable synthesis of enzymes involved in nucleotide metabolism during the G_1 phase, especially during late G_1/early S. RNA and protein synthesis occur continuously, although at varying rates during G_1, S, and G_2 phases, although DNA replication occurs only during S.

The strict regulation of nucleotide synthesis requires that certain mechanisms must be available to the cell to meet the requirements for the ribonucleotides and deoxyribonucleotide precursors at the time of increased RNA synthesis and DNA replication. To meet these needs, the cell responds by increasing the levels of specific enzymes involved with nucleotide formation during very specific periods of the cell cycle.

The enzymes involved in purine nucleotide synthesis and interconversions which are elevated during the S phase of the cell cycle are PRPP amidotransferase and IMP dehydrogenase. Adenylosuccinate synthetase and adenylosuccinase do not appear to increase.

The enzymes involved in pyrimidine nucleotide synthesis which are elevated during the S phase of the cell cycle include aspartate carbamoyltransferase, dihydroorotase, dihydroorotate dehydrogenase, orotate phosphoribosyltransferase, and CTP synthetase.

Many of the enzymes involved in the synthesis and interconversions of deoxyribonucleotides are also elevated during the S phase of the cell cycle. Included in these enzymes are ribonucleotide reductase, thymidine kinase, dCMP deaminase, thymidylate synthetase, and TMP kinase. The importance of the increases in these enzymes during S phase to DNA replication is worthy of further discussion with a specific example. As has been discussed previously, the deoxyribonucleotide pool is extremely small in "resting" cells (less than 1 μM). As a result of the increase in ribonucleotide reductase the levels of deoxyribonucleotides reach levels of 10–20 μM during DNA synthesis. However, this concentration would sustain DNA synthesis for only minutes, while complete DNA replication would require hours. Consequently, the levels of ribonucleotide reductase activity not only must increase but must be sustained during S phase in order to provide the necessary substrates for DNA synthesis.

If one looks at a population of cells as a whole (i.e., tissue) rather than as individual cells going through the cell cycle, it is observed that rapidly growing tissues such as regenerating liver, embryonic tissues, intestinal mucosal cells, and erythropoietic cells are geared toward DNA replication and RNA synthesis. These tissues will show elevated levels of those key enzymes involved with purine and pyrimidine nucleotide synthesis and interconversions and com-

plementary decreases in the levels of the enzymes that catalyze reactions in which these precursors are degraded. Of course these changes reflect the proportion of the cells in that tissue which are in S phase.

As a result of Weber's molecular correlation concept, an understanding has evolved of the biochemical changes that occur to satisfy the proliferative life-style of tumor cells. It has been determined that gene expression has been altered to result not only in quantitative changes in enzyme levels but also qualitative changes (isozyme shifts). As a result of careful experimental study, utilizing a series of liver and kidney tumors of varying growth rates, it has been possible to categorize these biochemical changes as (1) transformation-linked (meaning that all tumors regardless of growth rate show certain increased and certain decreased enzyme levels); (2) progression-linked (alterations that correlate with the growth rate of the tumor); and (3) coincidental alterations (not connected to the malignant state). As very limited examples, the levels of ribonucleotide reductase, thymidylate synthetase, and IMP dehydrogenase increase as a function of the tumor growth rate. PRPP amidotransferase, UDP kinase, and uridine kinase are examples of enzymes whose activity is increased in all tumors, whether they are slow-growing or the most rapidly growing tumors.

It is important to point out that, while certain of the enzymes are increased in both fast-growing normal tissue (e.g., embryonic and regenerating) and tumors, the total quantitative and qualitative patterns for normal and tumor tissue can easily be distinguished.

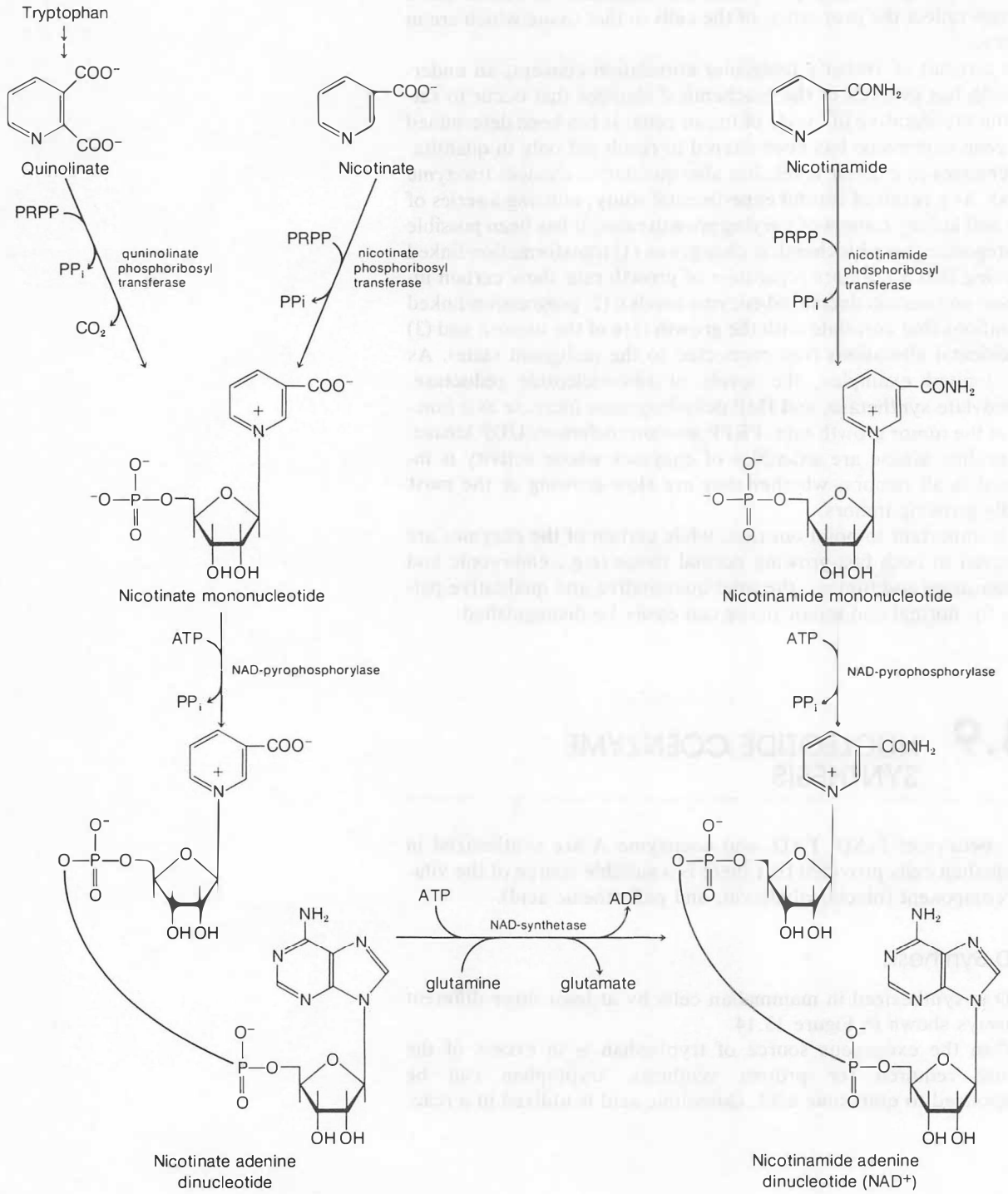
13.9 NUCLEOTIDE COENZYME SYNTHESIS

The coenzymes NAD, FAD, and coenzyme A are synthesized in mammalian cells provided that there is a suitable source of the vitamin component (niacin, riboflavin, and pantothenic acid).

NAD Synthesis

NAD is synthesized in mammalian cells by at least three different pathways shown in Figure 13.14.

When the exogenous source of tryptophan is in excess of the amount required for protein synthesis, tryptophan can be metabolized to quinolinic acid. Quinolinic acid is utilized in a reac-


Tryptophan
Quinolinate
Nicotinate
Nicotinamide
PRPP
PPi
CO2
quninolinate phosphoribosyl transferase
nicotinate phosphoribosyl transferase
nicotinamide phosphoribosyl transferase
Nicotinate mononucleotide
Nicotinamide mononucleotide
ATP
NAD-pyrophosphorylase
ADP
NAD-synthetase
glutamine
glutamate
Nicotinate adenine dinucleotide
Nicotinamide adenine dinucleotide (NAD+)

tion with PRPP to form nicotinate mononucleotide. The enzyme catalyzing this reaction, quinolinate phosphoribosyltransferase, is found only in the liver and kidney. Therefore, this pathway is specific for these tissues.

Nicotinate reacts with PRPP to form nicotinate mononucleotide. The enzyme catalyzing this reaction is nicotinate phosphoribosyltransferase and is widely distributed in various tissues. Nicotinate mononucleotide reacts with ATP to yield nicotinate adenine dinucleotide. The enzyme catalyzing this reaction is NAD-pyrophosphorylase and is widely distributed in various tissues. Nicotinate adenine dinucleotide reacts with glutamine with the hydrolysis of ATP to yield nicotinamide adenine dinucleotide (NAD). NAD-synthetase catalyzes this reaction.

Nicotinamide reacts with PRPP to give nicotinamide mononucleotide. The enzyme which catalyzes this reaction is nicotinamide phosphoribosyltransferase. The enzyme is specific for nicotinamide and is entirely distinct from nicotinate phosphoribosyltransferase. Nicotinamide mononucleotide reacts with ATP to yield NAD. The enzyme that catalyzes this reaction is the same enzyme that catalyzes the reaction between nicotinate mononucleotide and ATP.

In nucleated cells NAD-pyrophosphorylase is located exclusively in the nucleus of the cell. However, the erythrocyte is entirely capable of synthesizing NAD from nicotinate or nicotinamide.

The intracellular concentration of pyridine nucleotides is maintained at a constant level, implying a pathway that is tightly regulated. Nicotinamide phosphoribosyltransferase appears to be the regulated enzyme in NAD synthesis. NMN, NAD, NADP, and NADPH are strong inhibitors of nicotinamide mononucleotide synthesis. ATP stimulates this reaction, although it is not an absolute requirement. ATP lowers the K_m for PRPP 10-fold and nicotinamide 100-fold, while increasing the V_{max}. The resulting K_m for nicotinamide approaches the intracellular concentration of this compound. Nicotinate phosphoribosyltransferase does not appear to be regulated by the end products of the pathway utilizing nicotinate as the substrate.

NAD is synthesized also by the mitochondria. However, the relative importance of the mitochondrial pathway to the pathways just described is not known.

There is considerable turnover of NAD in cells. NAD is consumed by two distinct reactions. In one reaction NAD-glycohydrolase catalyzes the conversion of NAD to nicotinamide and adenosine

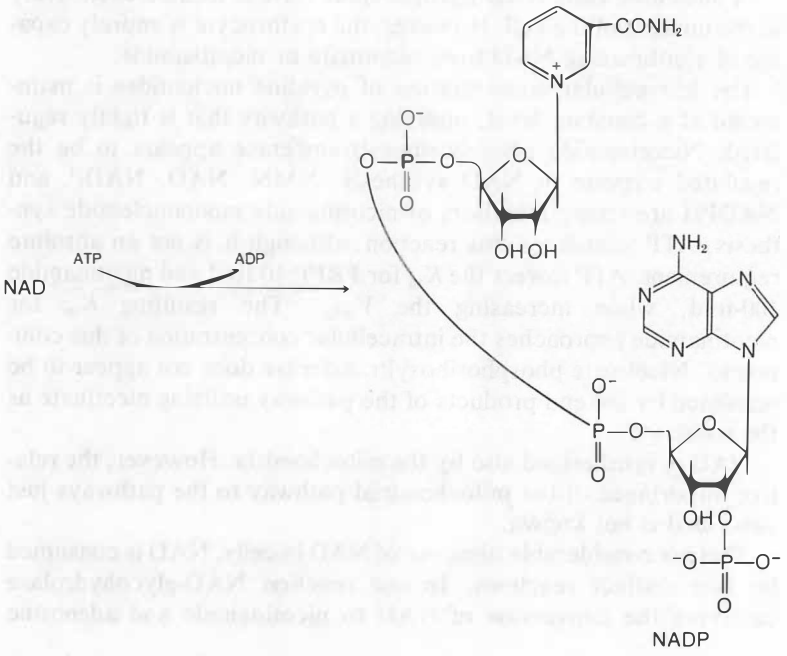
Figure 13.14
Pathways for NAD^+ synthesis.

diphosphoribose. This enzyme is located in microsomes. In the second, poly(ADP-ribose) synthetase catalyzes the polymerization of the ADP-ribose moiety of NAD onto nuclear proteins with the release of nicotinamide. This enzyme is found exclusively in the nucleus and the level of activity is highest during G_2 and lowest during S phase of the cell cycle. A possible role of polyADP ribosylation in cell regulation has been proposed.

There is no known enzymatic conversion of nicotinate directly to nicotinamide. Since only the liver and kidney can utilize tryptophan for NAD synthesis, the NAD glycohydrolase reaction can be thought to provide nicotinamide to extrahepatic tissues for NAD synthesis. In some tissues nicotinate is a more efficient precursor of NAD than is nicotinamide.

NADP Synthesis

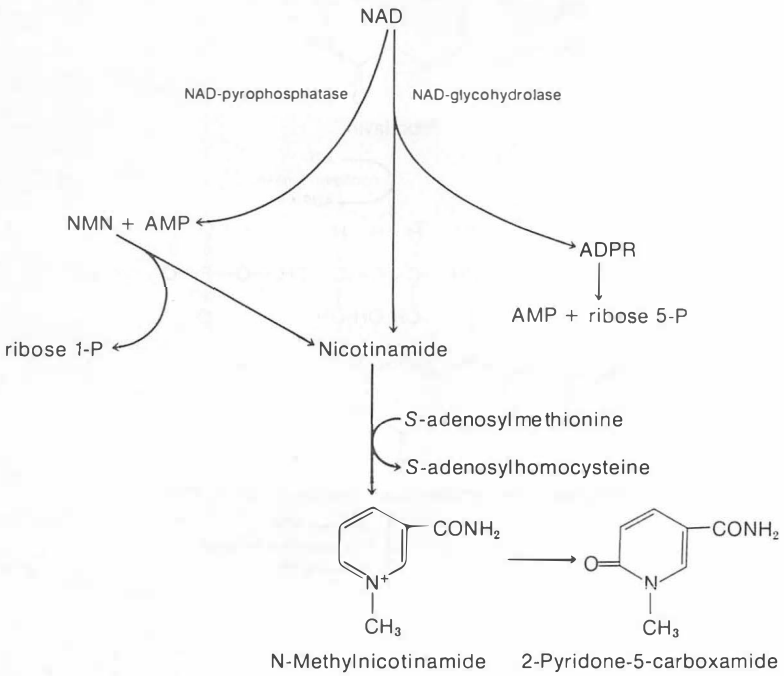
NAD is the immediate precursor of nicotinamide adenine dinucleotide phosphate (NADP). The reaction is catalyzed by NAD-kinase and is as follows:



NAD-kinase is found in the cytosol of the cells and NADPH, the reduced form, is a negative effector of this reaction.

Degradation of NAD

NAD can be degraded by two pathways, although the product is nicotinamide by either pathway.

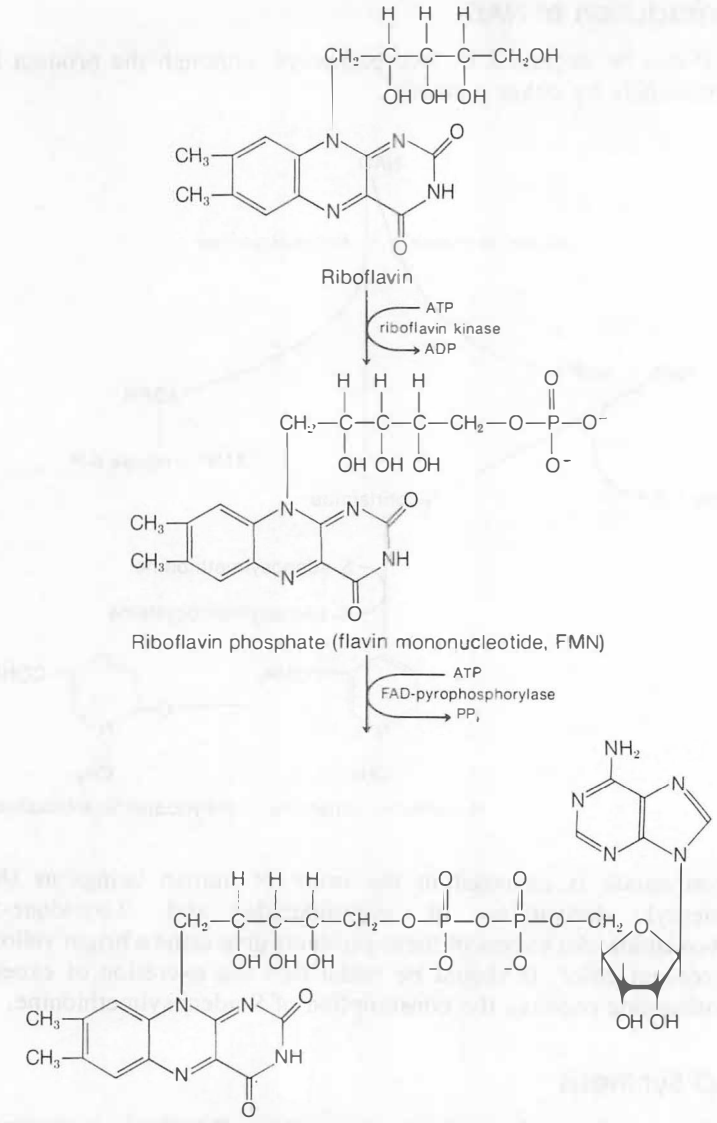


Nicotinamide is excreted in the urine of human beings as the N-methyl derivatives of nicotinamide and 2-pyridone-5-carboxamide. An excess of these products give urine a bright yellow fluorescent color. It should be noted that the excretion of excess nicotinamide requires the consumption of *S*-adenosylmethionine.

FAD Synthesis

FAD is synthesized in a two-step reaction. Riboflavin is required from an exogenous source because human beings cannot synthesize the isoalloxazine moiety. Riboflavin is phosphorylated by ATP in a reaction catalyzed by riboflavin kinase (flavokinase) to give riboflavin phosphate. Mg^{2+} ions are the preferred divalent cation required for this kinase. GTP will partially replace ATP as the phosphorylating agent. Flavokinase is found in the cytosol of the cell of a variety of tissues such as liver, kidney, brain, spleen, and heart. Many references are made to this compound as flavin mononucleotide

Figure 13.15
Synthesis of FAD.



(FMN), although this is not a true nucleotide. FMN then reacts with ATP to yield FAD in a reaction catalyzed by FAD pyrophosphorylase. This enzyme shows an absolute requirement for ATP since no other nucleoside triphosphate can substitute for ATP. Mg^{2+} ions are the required cations for this reaction. FAD-pyrophosphorylase activity has been reported to be located mainly in

mitochondria. The pathway for FAD synthesis is summarized in Figure 13.15.

Coenzyme A Synthesis

Coenzyme A is synthesized in human beings by a series of reactions, which has an absolute requirement for an exogenous source of pan-

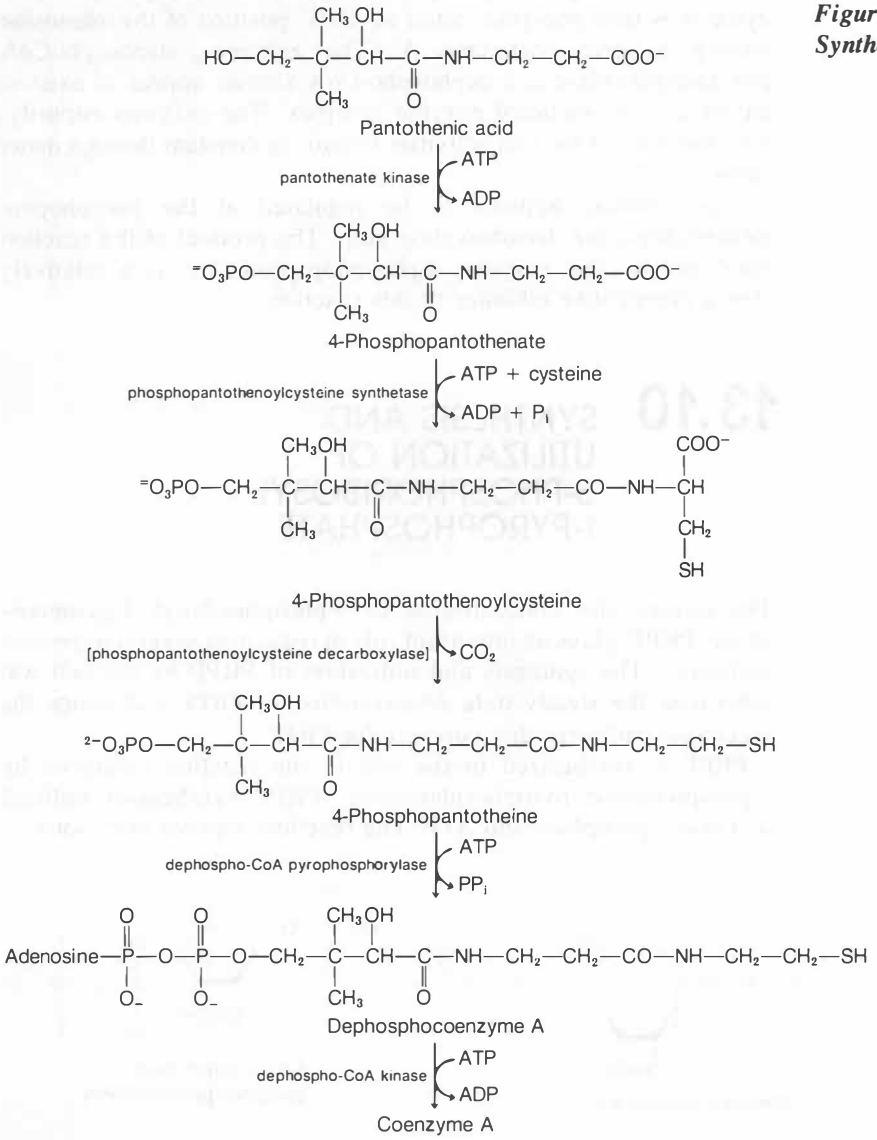


Figure 13.16
Synthesis of coenzyme A.

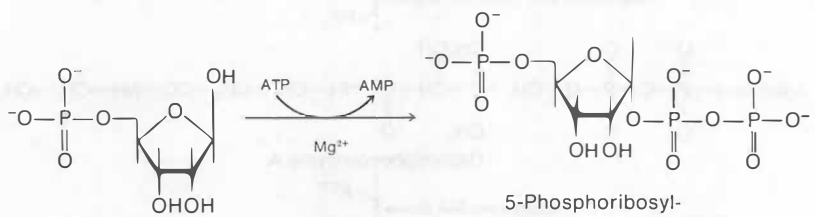
tothenic acid. The pathway for the synthesis of coenzyme A in mammalian cells is shown in Figure 13.16. Pantothenic acid is phosphorylated by ATP to give 4-phosphopantothenic acid. In the next reaction, cysteine is added to provide the —SH group, which will ultimately be the "business end" of CoA. The α-carboxyl group of cysteine is then removed from 4-phosphopantothenoyl-L-cysteine to yield 4-phosphopantotheine. In a pyrophosphorylase reaction, ATP is then added to give dephosphocoenzyme A. The dephosphocoenzyme A is then phosphorylated at the 3′ position of the adenosine moiety to give coenzyme A. The enzymes, dephosphoCoA pyrophosphorylase and dephospho-CoA kinase, appear to exist in nature as a bifunctional enzyme complex. The enzymes copurify, with the ratio of the two activities remaining constant through many steps.

This pathway appears to be regulated at the phosphopantothenoylcysteine decarboxylase step. The product of the reaction catalyzed by this enzyme, 4-phosphopantotheine, is a relatively strong competitive inhibitor of this reaction.

13.10 SYNTHESIS AND UTILIZATION OF 5-PHOSPHORIBOSYL 1-PYROPHOSPHATE

The intracellular concentration of 5-phosphoribosyl 1-pyrophosphate (PRPP) plays an important role in regulating several important pathways. The synthesis and utilization of PRPP by the cell will determine the steady-state concentration of PRPP and hence the metabolic pathways that compete for PRPP.

PRPP is synthesized in the cell in the reaction catalyzed by 5-phosphoribose pyrophosphokinase (PRPP synthetase) utilizing α-ribose 5-phosphate and ATP. The reaction requires Mg^{2+} ions.



Ribose 5-phosphate

5-Phosphoribosyl-pyrophosphate (PRPP)

The ribose 5-phosphate used in this reaction is generated from either glucose 6-phosphate metabolism via the hexose monophosphate shunt or ribose 1-phosphate (generated by phosphorylysis of nucleotides) via a phosphoribomutase reaction.

As expected for such a critical reaction, the formation of PRPP is regulated. The enzyme has an absolute requirement for inorganic phosphate ions. The velocity vs $[P_i]$ curve for PRPP-synthetase is sigmoidal and at the concentration of P_i normally found in the cell, the activity is markedly depressed because of the sigmoidal rather than hyperbolic curve.

The importance of the sigmoidal curve relative to inorganic phosphate in regulating PRPP formation, and hence the overproduction of uric acid, was shown in a "gouty" individual who had a marked increase in uric acid formation. Analysis of the patient's red cells for PRPP-synthetase revealed that this patient had a mutant form of the enzyme, which showed a hyperbolic v vs $[P_i]$ curve. At the intracellular concentration of P_i, the production of PRPP was greatly elevated. Presumably the increased levels of PRPP in other tissues lead to the overproduction of purine nucleotides and consequently to uric acid.

PRPP-synthetase is allosterically inhibited by nucleoside di- and triphosphates. 2,3-Diphosphoglyceric is also an inhibitor of PRPP formation, and this is probably important in red cell metabolism dealing with the salvage pathway.

PRPP formed in the cells is a required substrate for many key metabolic reactions depending on the cell type. These reactions and the pathways in which they are involved are as follows:

1. De novo purine nucleotide synthesis
   a. PRPP + glutamine $\longrightarrow$ 5-phosphoribosylamine
2. "Salvage" of purine bases
   a. PRPP + hypoxanthine (guanine) $\longrightarrow$ IMP (GMP)
   b. PRPP + adenine $\longrightarrow$ AMP
3. De novo pyrimidine nucleotide synthesis
   a. PRPP + orotate $\longrightarrow$ OMP
4. "Salvage" of pyrimidine bases
   a. PRPP + uracil $\longrightarrow$ UMP
5. NAD synthesis
   a. PRPP + nicotinic acid $\longrightarrow$ nicotinate mononucleotide

b. PRPP + nicotinamide ⟶ nicotinamide mononucleotide

c. PRPP + quinolinic acid ⟶ nicotinate mononucleotide

In the red cell, the major reactions in which PRPP would be consumed would be reactions 2a and 2b and 5a and 5b. In a rapidly growing tumor cell, PRPP would be consumed by all five reactions. The direction in which PRPP would be consumed would depend on several factors, including the relative K_m's of the competing enzymes for PRPP, the availability of the second substrate, and the concentration of the effector for the particular reaction. For example, if the concentration of AMP in the cell were high, adenine phosphoribosyltransferase would be inhibited and PRPP would be consumed via hypoxanthine-guanine phosphoribosyltransferase provided hypoxanthine or guanine were present in the cell and the IMP and GMP concentrations were low.

13.11 COMPOUNDS THAT INTERFERE WITH PURINE AND PYRIMIDINE NUCLEOTIDE METABOLISM

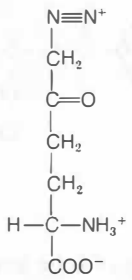
As has been discussed, the de novo synthesis of purine and pyrimidine nucleotides is critical to normal cell replication and function. The regulation of these pathways has also been shown to be important, since disease states arise from defects in these steps.

With this in mind, many compounds have been synthesized or isolated from plants, bacteria, fungi, and so on, that are directed at relatively specific metabolic sites involved with nucleotide synthesis or interconversion and these have been screened as potential antitumor agents. Several groups of drugs have emerged from these screens. The antifolates, glutamine antagonists, and antimetabolites are the major classes that have been identified.

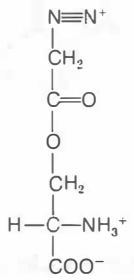
Glutamine Antagonists

Many reactions occur in mammalian cells in which glutamine serves as an amino donor. These amidation reactions are very critical for the synthesis of the purine ring de novo (N-3 and N-9), in the conversion of IMP to GMP, in the conversion of UTP to CTP and in the conversion of nicotinate adenine dinucleotide to NAD.

Compounds that inhibit these reactions are referred to as glutamine antagonists. Azaserine (*O*-diazoacetyl-L-serine) and 6-diazo-5-oxo-L-norleucine (DON), which were first isolated from cultures of *Streptomyces*, are very effective inhibitors of glutamine utilization. The structures of these glutamine antagonists are shown below:



6-Diazo-5-oxo-L-norleucine



Azaserine
(*O*-Diazoacetyl-L-serine)

Since azaserine and DON irreversibly inactivate the enzymes, glutamine will not reverse the effects of these two drugs. It would require that many types of metabolites such as guanine, cytidine, hypoxanthine (or adenine), and nicotinamide would have to be utilized to overcome the sites blocked by these glutamine antagonists. As expected from the multiple sites of inhibition at key steps, the glutamine antagonists are extremely toxic.

Antifolates

Antifolates are compounds that interfere with the formation of tetrahydrofolate from folic acid (Figure 13.17) and dihydrofolate. The decreased intracellular concentration of H_4folate results in the impaired formation of purine nucleotides via the de novo pathway and thymidylate via thymidylate synthetase.

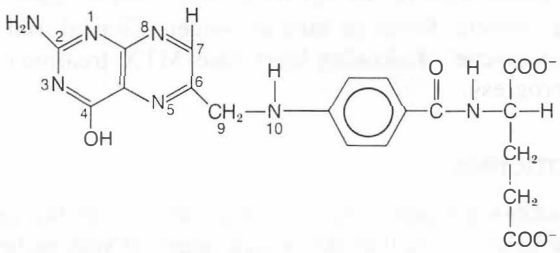
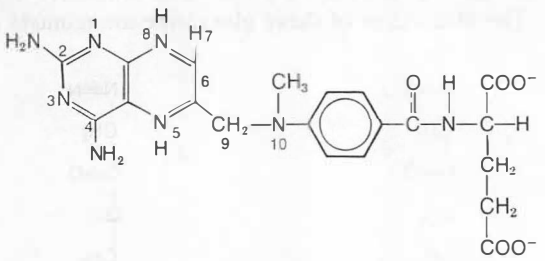


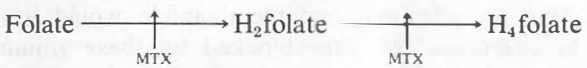
Figure 13.17
Structure of folic acid.

Methotrexate (MTX) is an example of an antifolate, which is currently in use in the treatment of various forms of cancer in human beings. The structure of methotrexate is shown below:



Methotrexate

Methotrexate is a close structural analog of folic acid. The differences in structure are at C-4, where an amino group replaces a hydroxyl group, and at N-10, where a methyl group replaces a hydrogen atom. The mode of action of MTX is quite specific. It inhibits dihydrofolate reductase with a K_i in the range of 10^{-10}M and thus inhibits the reactions:



When tumor cells in culture are treated with MTX, the cells die. However, the cytotoxic effects of MTX can be overcome by 5-formyl-H_4folate (citrovorum factor), thymidine and hypoxanthine. Some cell lines will respond to thymidine alone, but other cell lines will require both thymidine and hypoxanthine to rescue the cells from MTX treatment. The reversal by these compounds indicates that the direct effect of MTX on the H_4folate levels leads indirectly to the lack of thymidine and purine nucleotide formation.

Clinically, high-dose MTX therapy followed by "rescue" treatment with citrovorum factor has been successful in many situations in "curing" several forms of human cancer. Clinical trials utilizing "thymidine rescue" following high-dose MTX treatment are currently in progress.

Antimetabolites

Antimetabolites are generally structural analogs of the purine and pyrimidine bases or nucleosides which interfere with rather specific metabolic sites. Only a few of these will be discussed to show (1) the importance of the de novo pathways to normal cell metabolism; (2) that the regulation of these pathways occurs in vivo; (3) the concept

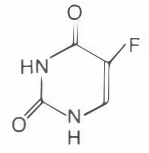
of the requirement for metabolic activation of the drugs; and (4) that the inactivation of these compounds can greatly influence their usefulness.

6-Mercaptopurine (*6-MP*) is a useful antitumor drug in man. The cytotoxic activity of this agent is related to the formation of 6-mercaptopurine ribonucleotide by the tumor cell. The tumor cell utilizing PRPP and HGPRTase converts 6-MP to its nucleotide form. 6-Mercaptopurineribonucleoside 5′-monophosphate accumulates in the cell and serves as a negative effector of PRPP-amidotransferase, the committed step in the de novo pathway. In addition, this nucleotide acts as an inhibitor of the conversion of IMP to GMP at the IMP-dehydrogenase step and IMP to AMP at the adenylosuccinate synthetase step. Since 6-mercaptopurine is a substrate for xanthine oxidase and is oxidized to 6-thiouric acid, allopurinol can be administered to inhibit the degradation of 6-MP and potentiate the antitumor properties of 6-mercaptopurine.

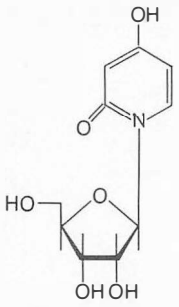
Adenine arabinoside (araA) has been used as an antitumor drug and as an antiviral agent in man. The effective metabolite of araA is its triphosphate, araATP. AraA, therefore, must be metabolized through araAMP and araADP. AraATP is an inhibitor of DNA-polymerase.

Cytosine arabinoside (araC) is currently in use in the treatment of several forms of human cancer. AraC, of itself, is not active. It must be converted to the nucleoside triphosphate (araCTP) to exert its cytotoxic effects. AraCTP is a potent competitive inhibitor, with respect to dCTP, of DNA polymerase. It has been shown that the effectiveness of araC as an antitumor drug correlates with the absolute level of araCTP achieved in the cell and the half-life of araCTP in the tumor cells.

5-Fluorouracil (FUra) is a pyrimidine analog, which of itself has no biological activity. FUra must be activated via the salvage pathways to the nucleotide level to exert its cytotoxic effects. FUra has been found to be converted to F-deoxyuridylate (FdUMP) and to FUTP. These two metabolites are the cytotoxic agents. FdUMP is a very potent and specific inhibitor of thymidylate synthetase. In the presence of thymidylate synthetase, H_4folate and FdUMP, a ternary complex is formed, which results in the covalent binding of FdUMP to thymidylate synthetase. FdUMP, therefore, causes what amounts to a "thymineless death" for the cells. The second metabolite, FUTP, is incorporated into the various RNA species of the cell. This incorporation of FUra into RNA has as its major consequence, the inhibition of the maturation of 45S precursor ribosomal RNA into the 28S and 18S species. The realization that the incorporation of FUra into RNA has serious effects on normal RNA metabolism and is a factor in the cytotoxicity of this agent has come only recently.



5-Fluorouracil



3-Deazauridine

The importance of these two metabolic sites to the cytotoxic action of this agent is confirmed by the fact that thymidine and uridine are both required to completely rescue the FUra-treated cells.

Deazauridine is a pyrimidine nucleoside analog that is an effective antitumor agent. This nucleoside must be "activated" to the di- and triphosphates to exert its cytotoxic effects. 3-Deazauridine-5'-triphosphate, an active form of this antimetabolite, is a potent inhibitor of CTP synthetase. DeazaUTP is a competitive inhibitor with respect to UTP of this enzyme. DeazaUDP is an inhibitor of ribonucleotide reductase activity. The net result of the inhibition at these sites is that the cells become deficient in cytidine and deoxycytidine nucleotides, causing inhibition of both RNA and DNA synthesis. In addition, CTP is required for normal phospholipid metabolism (e.g., CDP-choline).

Hydroxyurea

Tumor cells treated with hydroxyurea show a specific inhibition of DNA synthesis with little or no effect on RNA or protein synthesis. The metabolic basis for the specific inhibition of DNA synthesis is that hydroxyurea is an inhibitor of ribonucleotide reductase, blocking the reduction of all four nucleoside diphosphate substrates (CDP, UDP, GDP, and ADP). Toxicity to this drug results from the depletion of the deoxyribonucleoside triphosphates required for DNA replication.

Although this drug is specific for the inhibition of ribonucleotide reductase, clinically its use is limited due to its rapid rate of clearance and the high drug concentration that is required for effective inhibition of this enzyme.

The above drugs are only a few of the many compounds in clinical use, but serve as examples in which the knowledge of basic biochemical pathways and mechanisms lead to the generation of effective drugs. Another important point to be made regarding many of the antimetabolites used as drugs is that they must be activated to the nucleotide level to exert their cytotoxic effects. To become activated many of the cells' normal enzymes are utilized. The activation of 6-mercaptopurine to 6-mercaptopurine ribonucleotide requires the presence of HGPRTase activity and PRPP. Resistance of tumor cells to 6-mercaptopurine could result from a deficiency of HGPRTase activity. As the cells' enzymes can be utilized to convert the antimetabolites to the active form, the cells enzymes can also be utilized to inactivate these drugs. Two examples of this aspect are the drugs adenine arabinoside (araA) and cytosine arabinoside (araC). For araA to be an effective antiviral drug it must be converted to its nucleotide form, araATP. This takes place through the reactions

$$\text{araA} \longrightarrow \text{araAMP} \longrightarrow \text{araADP} \longrightarrow \text{araATP}$$

utilizing the cells' constitutive enzymes. On the other hand, araA is a substrate for adenosine deaminase. The product of this reaction, hypoxanthine arabinoside (araHx), is ineffective as a cytotoxic agent. Consequently, the effectiveness of araA as an antitumor or antiviral agent will depend, at least in part, on the ratio of the drug converted to araAMP and to araHx. Strategies using potent adenosine deaminase inhibitors, deoxycoformycin and erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA), have been developed in an attempt to potentiate the formation of araAMP and inhibit the production of araHx.

The activation of araC to araCTP depends on the presence of deoxycytidine kinase. Tumor cell resistance to araC could result from either a lack of deoxycytidine kinase or from an excess of deoxycytidine deaminase, which would inactivate araC to the araU derivative. Attempts to potentiate the activity of araC have involved combinations of drugs, which include inhibitors of deoxycytidine deaminase. Tetrahydrouridine is one such potent inhibitor of deoxycytidine deaminase. In fact, utilizing tetrahydrouridine, which itself has no effect on the cell, the concentration of araC required for cytotoxicity is markedly decreased.

These examples show the importance that the cells' enzymes play in drug action.

BIBLIOGRAPHY

General

Henderson, J. F., and Paterson, A. R. P. *Nucleotide metabolism, an introduction*. New York: Academic, 1973.

Stanbury, J. B., Wyngaarden, J. B., and Frederickson, D. S. (eds.). *The metabolic basis of inherited diseases*, 4th ed. New York: McGraw-Hill, 1978.

Weber, G. Enzymology of cancer cells. *N. Engl. J. Med.* 296:486, 1977.

Purine Metabolism

Henderson, J. F. *Regulation of purine biosynthesis*, ACS monograph 170. Washington, D.C.: American Chemical Society, 1972.

Wyngaarden, J. B. Regulation of purine biosynthesis and turnover. *Adv. Enz. Reg.* 14:25, 1976.

Pyrimidine Metabolism

Baliga, B. S., and Borek, E. Metabolism of thymine in tumor tissue: the origins of β-aminoisobutyric acid. *Adv. Enz. Reg.* 13:27, 1975.

Grunicke, J., Keller, H. J., Liersch, M., and Benaguid, A. New aspects of the mechanism and regulation of pyrimidine nucleotide metabolism. *Adv. Enz. Reg.* 12:397, 1974.

Traut, T. W., and Jones, M. E. Mammalian synthesis of UMP from orotate: the regulation of and conformers of complex U. *Adv. Enz. Reg.* 16:21, 1978.

Tremblay, G. C., Crandall, D. E., Knott, C. E., and Alfant, M. Orotic acid biosynthesis in rat liver: studies on the source of carbamyolphosphate. *Arch. Biochem. Biophys.* 178:264, 1977.

14

Metabolic Interrelationships

ROBERT A. HARRIS
DAVID W. CRABB

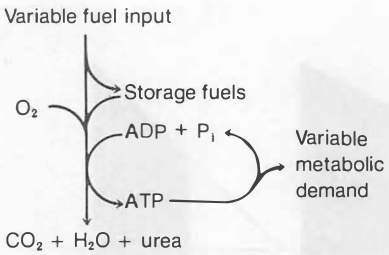


Figure 14.1
Human beings are able to use a variable fuel input to meet a variable metabolic demand.

CLIN. CORR. 14.1 OBESITY

Obesity is the most common nutritional problem in the United States and other affluent countries of the world. It causes a reduction in life span and is a risk factor in the development of diabetes mellitus, hypertension, osteoarthritis, gallstones, and cardiovascular diseases. Obesity is easy to explain—an obese person has eaten more than he or she required. The accumulation of massive amounts of body fat is not otherwise possible. Appreciation of this fact, however, is of little or no consolation to the obese individual. For unknown reasons the neural control of caloric intake to balance energy expenditure is abnormal. Rarely, obesity is secondary to a correctable disorder. Hypothyroidism causes puffiness and some weight gain but is hardly ever the cause of obesity. Hypercortisolism (Cushing's syndrome) results in central obesity

14.1 OVERVIEW

In this chapter the interdependence of the metabolic processes of the major tissues of the body in physiological situations will be stressed. Not all of the major metabolic pathways and processes of the body operate in every tissue at any given time. Given the nutritional and hormonal status of a patient, we need to be able to say, at least qualitatively, which of the major metabolic pathways of the body are functional and how these pathways relate to one another.

The metabolic processes with which we are concerned in this chapter are glycogenesis, glycogenolysis, gluconeogenesis, glycolysis, fatty acid synthesis, fatty acid oxidation, citric acid cycle activity, ketogenesis, amino acid oxidation, protein synthesis, proteolysis, and urea synthesis. It is important to know (1) which tissues are most active in these various processes, (2) when these processes are most and least active, and (3) how these processes are controlled and coordinated in different metabolic states.

The best way to gain an understanding of the relationships of the major metabolic pathways to one another is to become familiar with the changes in metabolism that occur during the starve–feed cycle. As shown in Figure 14.1, the starve–feed cycle allows a variable fuel consumption to meet a variable metabolic demand. Starve is a poor word choice in this case; fast is what we mean, but the phrase fast–feed cycle brings to mind lunch at McDonald's rather than what we are trying to express. Feed refers to the intake of meals (the variable fuel input) after which we store the fuel (in the form of glycogen and fat) to be used to meet our metabolic demand while we fast. Note the participation of an ATP cycle within the starve–feed cycle (Figure 14.1). ATP functions as the energy-transferring agent in the starve–feed cycle, being like money to the cell—the payoff to the cell for going to the trouble of burning fuels to CO_2 and H_2O.

Humans have the capacity to consume food at a rate some 100 times greater than our basal caloric requirements. This allows us to survive from meal to meal without nibbling continuously between meals. We thus store the calories as glycogen and fat and consume them as needed. Unfortunately, an almost unlimited capacity to consume food is matched by an almost unlimited capacity to store it as fat. Obesity, a big problem in this country, is the consequence of excess consumption (Clin. Corr. 14.1), whereas other forms of malnutrition are more prevalent in the emerging, less affluent countries of the world (Clin. Corr. 14.2). Every day represents a series of starve–feed cycles for a person. It balances out quite nicely for most, that is, fuel consumption equals fuel utilization, but others are not so fortunate. The regulation of food consumption is complex and

not well understood. The tight control needed is indicated by the calculation that two extra pats of butter (~100 cal) per day over caloric expenditures results in 10 lb weight gain per year.

(face, trunk) with wasting of the limbs, hypertension, glucose intolerance, and stretching of subcutaneous connective tissue. Insulin-secreting tumors of the β cell (insulinoma) may cause weight gain because the hypoglycemia resulting from inappropriate insulin secretion causes hunger. Even more rare are tumors, vascular accidents, or maldevelopment of the nervous system hunger control centers in the hypothalamus. The most common "gland" problem causing obesity is, of course, overactivity of the salivary glands.

In the most common type of obesity, the number of adipocytes of the body does not increase, they just get larger as they become engorged with triacylglycerols. If obesity develops before puberty, however, an actual increase in the number of adipocytes can also occur. In the latter case both hyperplasia (increase in cell number) and hypertrophy (increase in cell size) are contributing factors to the magnitude of the obesity.

The only effective treatment of obesity is reduction in the ingestion, absorption, or utilization of calories. Practically, this means dieting.

14.2 STARVE-FEED CYCLE

Well-Fed State

Figure 14.2 shows what happens to glucose, amino acids, and fat obtained from food in the gut. Note the different route by which fat makes its way into the bloodstream. Glucose and amino acids pass directly into the blood from the intestinal epithelial cells and are presented to the liver by way of the portal vein. Fat, as chylomicra, passes from the intestinal epithelial cells into lymphatics, which drain the intestine. The lymphatics lead to the thoracic duct, which, by way of the subclavian vein, delivers chylomicra to the blood at a site of rapid blood flow. The latter provides for rapid distribution of the chylomicra and prevents coalescence of the fat particles.

Liver is the first tissue to have the opportunity to utilize dietary glucose. After penetrating the plasma membrane of this tissue, glucose can be converted into glycogen by glycogenesis, into pyruvate and lactate by glycolysis, or used in the pentose phosphate pathway for the generation of NADPH for reductive synthetic processes. Pyruvate formed from glucose can be oxidized to acetyl CoA, which, in turn, can be converted into fat by the process of lipogenesis or oxidized to CO_2 and water by the TCA cycle. Much of the glucose coming from the intestine escapes the liver and circulates to other tissues of the body. The brain is a major user of blood glucose, being almost solely dependent upon the catabolism of this substrate to CO_2 and water for its production of ATP. Other major users of glucose include red blood cells, which can only convert glucose to lactate and pyruvate, and the adipose tissue, which converts it into fat. Muscle in the well-fed state has the capacity to utilize glucose, converting it to glycogen or using it in the glycolytic and TCA cycle pathways. A number of tissues produce, by way of glycolysis, lactate and pyruvate from circulating glucose. The pyruvate and lactate circulate in the blood to the tissues that are active metabolizers of these substrates. In the well-fed state, liver and adipose tissue are avid users of lactate and pyruvate for the process of lipogenesis, although in the human most fat is synthesized in the liver rather than in the adipose tissue. In the very well-fed state, the liver utilizes glucose and does not engage in gluco-

CLIN. CORR. 14.2 PROTEIN MALNUTRITION

Protein malnutrition is the most important and widespread nutritional problem among young children in the world today. The clinical syndrome, called kwashiorkor, occurs mainly in children 1 to 3 years of age, often being precipitated by weaning an infant from breast milk onto a starchy, protein-poor diet. The name originated in Ghana, meaning "the sickness of the older child when the next baby is born." Common in many developing countries, it is a consequence of feeding the child a diet adequate in calories but deficient in protein. It may become clinically manifest when protein requirements are increased by infection, for example, malaria, helminth infection, or gastroenteritis. The syndrome is characterized by poor growth, low plasma protein levels, muscle wasting, edema, diarrhea, and increased susceptibility to infection. The presence of subcutaneous fat clearly differentiates it from caloric starvation. The lack of dietary amino acids results in diminished protein synthesis in all tissues. The liver becomes enlarged and infiltrated with fat, reflecting the need for hepatic protein synthesis for the formation and release of lipoproteins from this tissue. In addition, protein malnutrition impairs the function of the gut, resulting in malabsorption of calories, protein, and vitamins, and accelerates the disease. The consequences of the disease include a permanent stunting of physical growth and poor intellectual and psychological development.

neogenesis. Thus, the Cori cycle, which involves conversion of glucose to lactate in the peripheral tissues followed by conversion of lactate back to glucose in the liver, is interrupted in the well-fed state.

Liver is also the first tissue to have the opportunity to remove absorbed amino acids from the blood (Figure 14.2). As a general rule the liver lets most of each amino acid pass through, unless the concentration of the amino acid is unusually high. This is especially important for the essential amino acids, needed by all tissues of the body for protein synthesis. The liver can catabolize amino acids, but the K_m values of many of the enzymes involved for amino acids are high, allowing the amino acids to be present in excess before significant catabolism can occur. In contrast, the tRNA charging enzymes involved in the synthesis of proteins have much lower K_m values for amino acids. This ensures that as long as all the amino acids are present, protein synthesis can occur as needed for growth and replacement synthesis purposes. Amino acids that are catabolized in the liver can be oxidized completely to CO_2 and water, or the intermediates generated can be used as substrates for lipogenesis, ketogenesis, or gluconeogenesis. Thus, excess amino acids not needed for protein synthesis end up being converted to ketone bodies, glucose, or fat, with the amino nitrogen being converted to urea. Amino acids that escape the liver can be used for protein synthesis in other tissues. Skeletal muscle and heart muscle have a high capacity for amino acid transamination and oxidation of the resulting α-keto acids to CO_2 and water. The branched chain amino acids (leucine, isoleucine, and valine) are handled in an interesting manner. The liver has low capacity for the transamination of these amino acids but considerable capacity for oxidative decarboxylation of the α-keto acids derived from these amino acids. The opposite is true for skeletal muscle, that is, this tissue has considerable capacity for transamination but is relatively deficient in the enzymes responsible for subsequent catabolism. The consequence is that much of the transamination occurs in peripheral tissues like skeletal muscle, the α-keto acids escape into the blood, and the liver oxidizes the α-keto acid. Branched-chain amino acids are a major source of nitrogen for the production of alanine from pyruvate in muscle.

When considering fat delivery to the tissues, we must differentiate carefully between endogenous and exogenous fat (Figure 14.2). We have already discussed above that glucose, lactate, pyruvate, and amino acids can be used to support hepatic lipogenesis. The fat formed from these substrates is released from the liver in the form of VLDL for transport primarily to adipose tissue. Dietary fat is delivered to the bloodstream as chylomicra. Both chylomicra and VLDL circulate in the blood until they are acted upon by a special extracel-

lular enzyme attached to the capillary wall of many tissues. This enzyme, lipoprotein lipase, is particularly abundant on the surface of the capillaries in adipose tissue. It acts on both the VLDL coming from the liver and the chylomicra coming from the gut, liberating fatty acids by hydrolytic cleavage of the triacylglycerols. The fatty acids are then taken up by the adipocytes, reesterified with α-glycerol phosphate for the eventual formation of triacylglycerols, and stored as large fat droplets within these cells.

The last thing to be noted about Figure 14.2 is that the β cells of

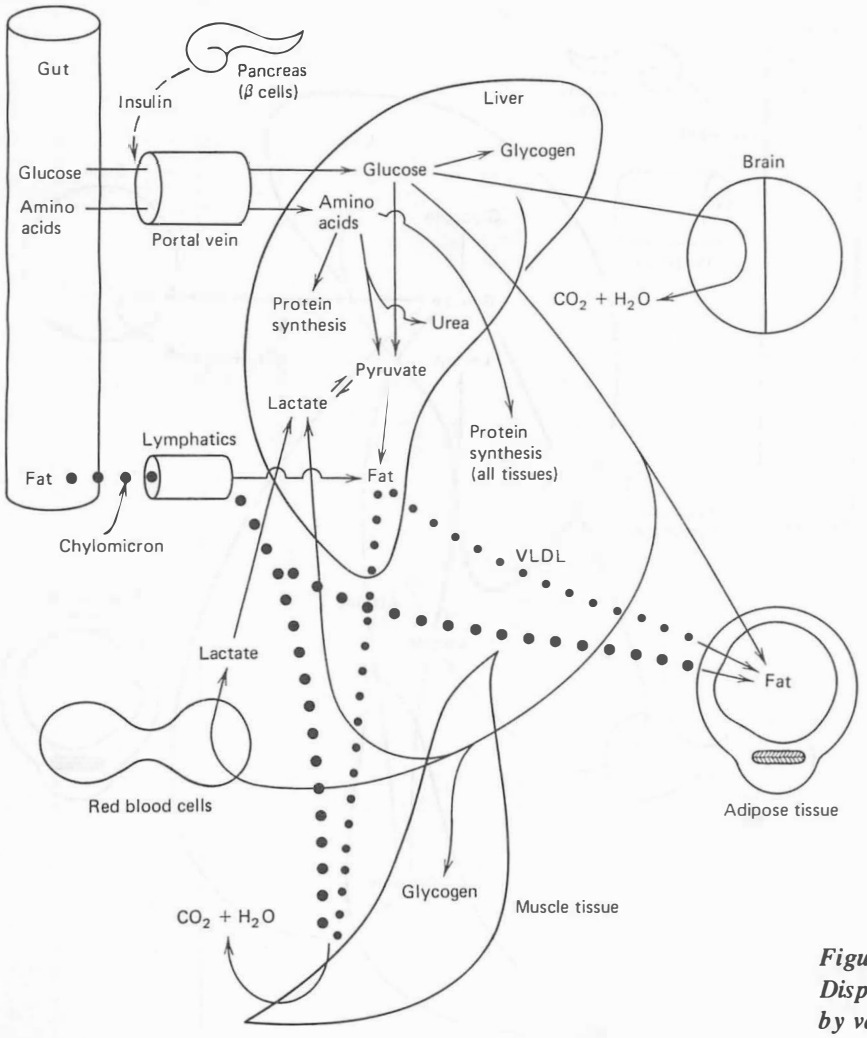


Figure 14.2
Disposition of glucose, amino acids, and fat by various tissues in the well-fed state.

the pancreas are very responsive to elevated blood glucose levels, releasing insulin to facilitate the utilization of glucose by liver, muscle, and adipose tissue. Insulin is extremely important for the proper metabolism of all foodstuffs. Indeed, just about everything that happens in the well-fed state to glucose, amino acids, and fat, is known to be dependent either directly or indirectly upon a high insulin : glucagon ratio, as discussed in some detail in Section 14.3.

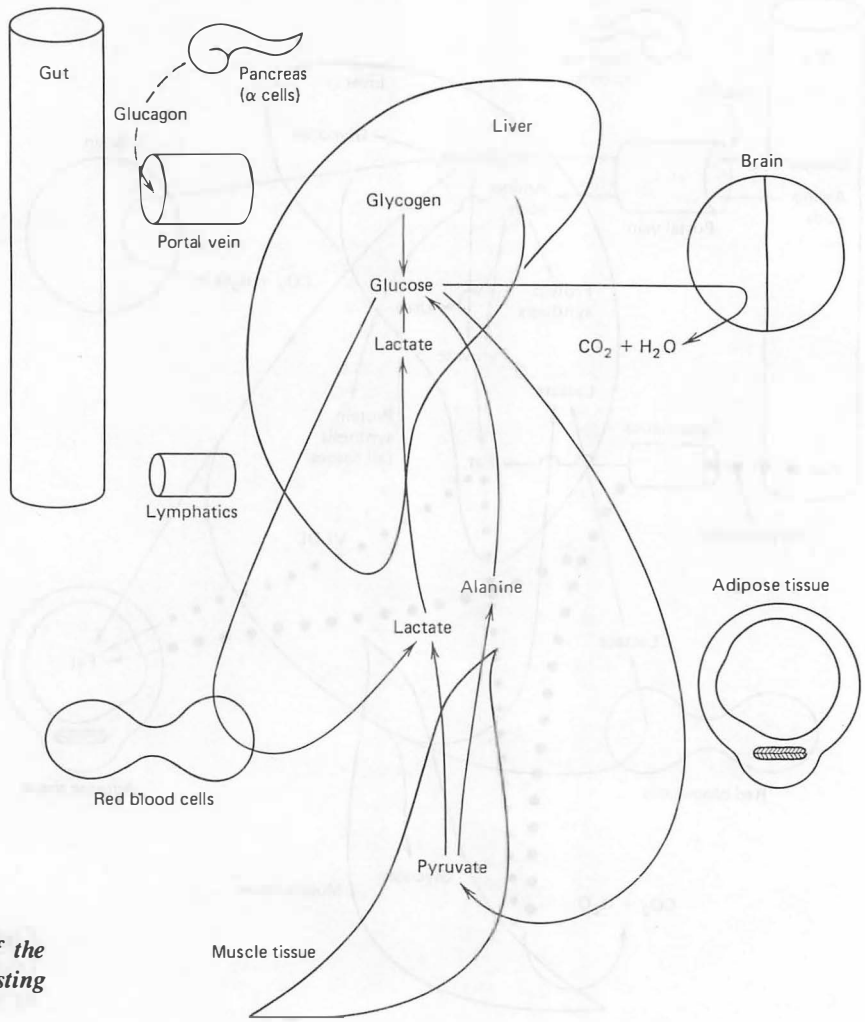


Figure 14.3
Relationships between the metabolism of the major tissues of the body in the early fasting state.

Early Fasting State

Figure 14.3 shows what happens after fuel stops coming in from the gut, that is, in early fasting. Glycogenolysis of hepatic glycogen for the maintenance of blood glucose is very important during this transitional period. Lipogenesis is curtailed and the lactate, pyruvate, and amino acids that were being used to support that process are diverted into the formation of glucose. Thus the Cori cycle is shown as part of Figure 14.3, glucose being produced from lactate by the liver and then being converted back to lactate by glycolysis in peripheral tissues such as red blood cells. The alanine cycle, in which carbon returns to the liver in the form of alanine rather than lactate, also becomes increasingly important as a mechanism for maintaining blood glucose levels. Because less is available from the gut, and the blood levels are in general lower, the catabolism of amino acids as a source of energy by tissues is greatly diminished in the early fasting condition.

Fasting State

Figure 14.4 shows what happens as we move into the actual fasting state. No fuel is coming in from the gut at all in this situation, and no glycogen is left in the liver to help maintain blood glucose, which is completely dependent upon hepatic gluconeogenesis, primarily from lactate, glycerol, and alanine. Again the Cori and the alanine cycles play important roles in supplying glucose for tissues that are dependent upon this substrate for the production of ATP. It must be noted, however, that the Cori cycle and the alanine cycle do not provide carbon for net synthesis of glucose. In these cycles glucose formed by the liver just replaces that which was converted to lactate by the peripheral tissues. There is no net glucose synthesis. The brain oxidizes glucose completely to CO_2 and water and probably does not participate to a significant extent in either cycle. Hence net glucose synthesis from some source of carbon is mandatory in this situation. Fatty acids cannot be used for the net synthesis of glucose, there being no pathway in animals for such a conversion. Specifically, acetyl CoA obtained by fatty acid catabolism cannot be converted to three-carbon intermediates of gluconeogenesis. Glycerol, obtained from lipolysis of fat in adipose tissue, functions as an important substrate for glucose synthesis in the fasted state. However, protein of the various tissues of the body, especially skeletal muscle, supplies most of the carbon needed for net glucose synthesis under such conditions. The sequence of events can be pictured as follows. Proteins are hydrolyzed within muscle cells (proteolysis) to produce amino acids. A portion of these amino acids escapes into the blood and circulates to other tissues. Somewhat surprisingly, most of the

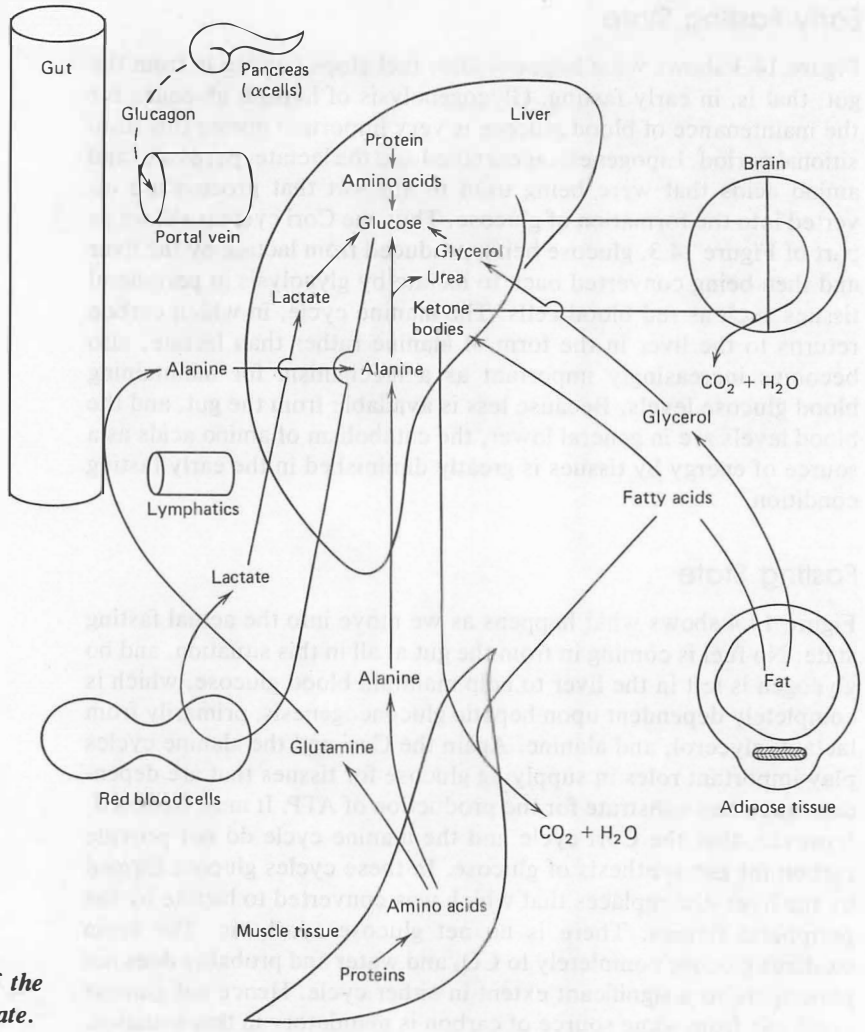


Figure 14.4
Relationships between the metabolism of the major tissues of the body in the fasting state.

amino acids do not escape but are partially metabolized within the muscle cells, where proteolysis took place. Only three amino acids—alanine, glutamine, and glycine—are released in large amounts. What appears to happen is that the other amino acids are metabolized by their various catabolic pathways to give intermediates (pyruvate and α-ketoglutarate), which can yield alanine and glutamine. These amino acids are then released into the blood, from which they can be removed by the liver for net glucose formation. There is evidence that much of the glutamine is converted into alanine by cells of the kidney and the intestinal epithelium. Glutamine is oxidized partially in these cells, supplying energy to

meet part of the metabolic demand of these tissues, and the carbon and amino groups left over are released back into the bloodstream in part as alanine and NH_4^+. (The exact pathway for the latter has not been established, but probably involves the net formation of citric acid cycle intermediates from glutamine and the stepwise conversion of oxalacetate to phospho*enol*pyruvate, phospho*enol*pyruvate to pyruvate, and pyruvate to alanine.) Hence, alanine is quantitatively the most important gluconeogenic amino acid to reach the liver in the fasting state. Glycine released from muscle is transformed in part into serine by the kidneys. Serine released by the kidneys into the blood is subsequently converted into glucose by the liver.

The adipose tissue also becomes very much involved in the fasting state. Because of the low insulin : glucagon ratio existing in this condition, lipolysis within this tissue is greatly activated. This results in elevated blood levels of fatty acids, which can be used as alternative fuels to glucose by many tissues. Heart and skeletal muscle actually prefer to use fatty acids over glucose, the oxidation of fatty acids inhibiting glycolysis. The brain, on the other hand, does not use fatty acids as a substrate, these compounds penetrating very poorly the blood–brain barrier. The reason for this is not clear, but we can at least imagine that it is important to keep powerful detergents like fatty acids away from a tissue as delicate as brain. Fatty acids play a very important role in liver, being oxidized by the β oxidation pathway for the provision of most of the ATP needed to support the energy requirements of gluconeogenesis. Very little of the acetyl CoA generated by β oxidation in the liver is oxidized completely to CO_2 and water by the TCA cycle. Since a low rate of TCA cycle activity occurs during gluconeogenesis, the acetyl CoA formed from fatty acids is converted into ketone bodies by liver mitochondria under these conditions. The ketone bodies (acetoacetate and β-hydroxybutyrate) are released into the blood to be used as a source of energy by several different tissues. Like fatty acids, ketone bodies are preferred by many tissues over glucose, the presence of ketone bodies and fatty acids having a sparing effect upon blood glucose so glucose can be available for use by the brain. Ketone bodies, in contrast to fatty acids, penetrate the blood–brain barrier. Once their blood concentration is high enough, ketone bodies function as a good alternative fuel for the brain. They are unable, however, to replace the need for glucose by the brain completely. Ketone bodies do decrease brain glucose utilization and, as if anticipating this effect, also suppress proteolysis in skeletal muscle and decrease to a certain extent muscle wasting, which inevitably occurs during starvation. As long as ketone body levels are maintained at a high level by hepatic β oxidation, there is less need for glucose, less need for gluconeogenic amino acids, and less need for using up precious muscle tissue by proteolysis.

CLIN. CORR. 14.3 REYE'S SYNDROME

Reye's syndrome presents in children as encephalopathy with evidence of hepatic dysfunction. It is characterized by hypoglycemia, coma, elevated blood ammonia levels, and fatty liver. It is relatively uncommon and the cause uncertain, although patients typically are recovering from a recent viral infection. The liver shows engorgement with triacylglycerols and morphologic changes in the mitochondria, which may account for the decreased capacity for gluconeogenesis, fatty acid oxidation, and urea synthesis. There is little liver inflammation or cell death, and recovery can be complete.

The working relationship between liver, muscle, and adipose tissue in supplying glucose for the brain should be noted in Figure 14.4. The liver functions to synthesize the glucose, muscle supplies the substrate (alanine), and adipose tissue supplies the ATP (via fatty acid oxidation) needed for hepatic gluconeogenesis. This interaction is dependent upon and favored by a low insulin : glucagon ratio—opposite to the ratio favoring those processes that are characteristic of the well-fed state. Glucose levels are lower in the fasting condition, preventing the release of insulin but favoring the release of glucagon from the pancreas to give a much lower insulin : glucagon ratio than in the well-fed state.

Caloric Homeostasis

The major tissues of the body work closely together in an effort to maintain a constant availability of oxidizable fuels in the blood. This is termed caloric homeostasis, which, as illustrated in Table 14.1, means that regardless of whether a person is in the well-fed state, fasting, or starving to death, the blood level of ATP equivalent fuel does not fall below certain limits. This is accomplished, as discussed above, by mobilizing the energy stores of the body so that all tissues will have ready access to oxidizable fuels to supply the ATP needed to meet their metabolic demands. The changes in insulin : glucagon ratio, discussed above and shown in Table 14.1, are crucial to the proper maintenance of caloric homeostasis. Note that blood glucose levels are controlled within very tight limits, whereas fatty acid concentrations in the blood can vary by an order of magnitude and ketone bodies by two orders of magnitude. The fact that glucose is maintained within tighter limits is again related to the absolute need of the brain for this substrate. If the blood glucose level falls too low (<1.5 mM), coma results from lack of ATP production, and death will follow shortly unless the situation can be rapidly corrected (see Clin. Corr. 14.3). On the other hand, hyperglycemia is to be avoided because of the possibility of developing hyperosmolar, hyperglycemic coma (see Clin. Corr. 14.4). Hyperglycemia also leads to the glycosylation of a number of proteins, postulated to be a more insidious complication of excessive concentrations of glucose (see Clin. Corr. 14.5).

Energy Requirements and Reserves of Man

The average person leading a sedentary life, such as a portly professor of biochemistry, consumes daily about 200 g of carbohydrate, 70 g of protein, 60 g of fat, and, during the academic year, 100 g of ethanol and an occasional graduate student. As shown in Table 14.2,

Table 14.1 Substrate and Hormone Levels in the Blood of the Well-fed, Fasting, and Starving Man[a]

Hormone or Substrate (units)	*Very Well-Fed*	*Postabsorptive 12 hours*	*Fasted 3 days*	*Starved 5 weeks*
Insulin (μU/ml)	40	15	8	6
Glucagon (pg/ml)	80	100	150	120
Insulin : glucagon ratio (μU/pg)	0.50	0.15	0.05	0.05
Glucose (mM)	6.1	4.8	3.8	3.6
Fatty acids (mM)	0.14	0.6	1.2	1.4
Acetoacetate (mM)	0.04	0.05	0.4	1.3
β-Hydroxybutyrate (mM)	0.03	0.10	1.4	6.0
Lactate (mM)	2.5	0.7	0.7	0.6
Pyruvate (mM)	0.25	0.06	0.04	0.03
Alanine (mM)	0.8	0.3	0.3	0.1
ATP equivalents (mM)	313	290	380	537

[a] Data are for normal weight subjects except for the 5-week starvation values, which are from obese subjects undergoing therapeutic starvation. ATP equivalents were calculated on the basis of the ATP yield expected on complete oxidation of each substrate to CO_2 and H_2O: 38 molecules of ATP for each molecule of glucose; 144 for the average fatty acid (oleate); 23 for acetoacetate; 26 for β-hydroxybutyrate; 18 for lactate; 15 for pyruvate; and 13 (corrected for urea formation) for alanine. Taken in part from: Ruderman, N. B., Aoki, T. T., and Cahill, G. F., Jr. Gluconeogenesis and its disorders in man, in R. W. Hanson, and M. A. Mehlman (eds.), *Gluconeogenesis, its regulation in mammalian species*. New York: Wiley, 1976, p. 515.

the energy reserves of even an average person are indeed considerable. We tend to emphasize the details of glycogen metabolism, and the ability to mobilize glycogen rapidly is indeed very important. Table 14.2 makes the point, however, that our glycogen reserves are minuscule with respect to our fat reserves. Obesity is characterized by further increases in fat reserves with little or no significant increase in glycogen reserves. The fat stores of obese subjects can weigh as much as 80 kg, adding another 585,000 kcal to their energy reserves. Protein is listed in Table 14.2 as an energy reserve, and, in the sense that it can be used to provide substrate for gluconeogenesis or amino acid oxidation, protein is an energy reserve. On the other hand, protein is not inert like stored fat and glycogen. It corresponds to the proteins that make up the muscles that allow us to move around, and to the enzymes that carry out all the catabolic and anabolic processes. Hence it is not as dispensable as fat and glycogen and is given up by the body more reluctantly.

CLIN. CORR. 14.4 HYPEROSMOLAR, HYPERGLYCEMIC COMA

Uncontrolled hyperglycemia (glucose levels 500 mg % and higher) in diabetes is life-threatening apart from ketoacidosis. It may be precipitated by infection, dietary indiscretion, or missing doses of insulin. High levels of glucose result in an osmotic gradient between cells and blood. Under this osmotic pressure water moves out of cells, shrinking them, and resulting in generalized cellular malfunction, particularly in the central nervous system. Glucose is lost into the urine, where it also is osmotically active, increasing renal losses of water and electrolytes. The patient is therefore depleted of fluid both intra- and extracellularly. This impairment of blood volume can lead to decreased tissue perfusion, lactic acidosis, hypotension, and death. Treatment must not reduce blood glucose too rapidly because the intracellular glucose concentration is high in these patients, and this glucose can not reequilibrate with the blood instantaneously. If blood glucose falls below cellular glucose, the osmotic gradient is reversed, water flows *into* the cells, and cell swelling occurs. This may worsen nervous system dysfunction further.

CLIN. CORR. 14.5 HYPERGLYCEMIA AND GLYCOSYLATION

Glycosylation of enzymes is known to cause changes in their activity, solubility, and susceptibility to degradation. In the case of hemoglobin A, glycosylation occurs by a nonenzymatic reaction between glucose and the amino-terminal valine of the β chain. A Schiff's base between glucose and valine forms, followed by a rearrangement of the molecule to give a 1-deoxyfructose molecule attached to the valine. The reaction is favored by high glucose levels and the resulting protein, called hemoglobin A_{Ic}, is a good index of how high a person's blood-glucose concentration has been over the previous several weeks. The concentration of this protein increases substantially in the red blood cells of an uncontrolled diabetic.

It has been proposed that increased glycosylation of proteins resulting from hyperglycemia may contribute to the medical complications caused by diabetes, for example, coronary heart disease, retinopathy, nephropathy, and neuropathy. High blood glucose concentrations increase the intracellular glucose concentration of tissues that do not require insulin for glucose uptake. The characteristic thickening of the basement membranes of capillaries of the diabetic patient may be due to increased synthesis of glycoproteins by this mechanism. Likewise, increased glycosylation of the lens protein α-crystallin may contribute to the development of diabetic cataracts.

Table 14.2 The Energy Reserves of Man[a]

Stored Fuel	*Tissue*	*Fuel Reserves in Units of*	
		Grams	*Kilocalories*
Glycogen	Liver	70	280
Glycogen	Muscle	120	480
Glucose	Body fluids	20	80
Fat	Adipose	15,000	135,000
Protein	Muscle	6,000	24,000

[a] The data are for a normal subject weighing 70 kg. Carbohydrate is taken to be worth 4 kcal/g; fat 9 kcal/g; protein 4 kcal/g.

The Five Phases of Glucose Homeostasis

Figure 14.5 comes from the work of Cahill and colleagues with obese patients undergoing long-term starvation for therapeutic purposes. It illustrates the effects of starvation on those processes that are used by the tissues of the body to maintain caloric homeostasis. For convenience of discussion, the time period involved has been divided into five phases. Phase I simply refers to the well-fed state, in which the glucose being used by the body is provided by that coming from dietary carbohydrate. Once this supply is exhausted, glycogenolysis in the liver maintains blood glucose levels during phase II. As this supply of glucose starts to dwindle, gluconeogenesis within the liver from lactate, glycerol, and alanine becomes increasingly important until, in phase III, it is the major source of blood glucose. Note that all of these changes occur within just 20 or so hours of fasting, depending of course on how well fed the individual was prior to the fast, how much hepatic glycogen was present, and the sort of physical activity occurring during the fast. Several days of fasting move man into phase IV, where the dependence upon gluconeogenesis actually decreases. The explanation for this surprising phenomenon, already discussed above, is that ketone bodies have accumulated to concentrations that are high enough for them to enter the brain and meet some of the energy needs of this tissue. Phase V occurs after very prolonged starvation of extremely obese individuals. It is characterized by even less dependence of the body upon gluconeogenesis, the energy needs of every tissue containing mitochondria being met to an even greater extent by either fatty acid or ketone body oxidation.

As long as ketone body concentrations are high, proteolysis will be somewhat restricted, and conservation of muscle proteins and

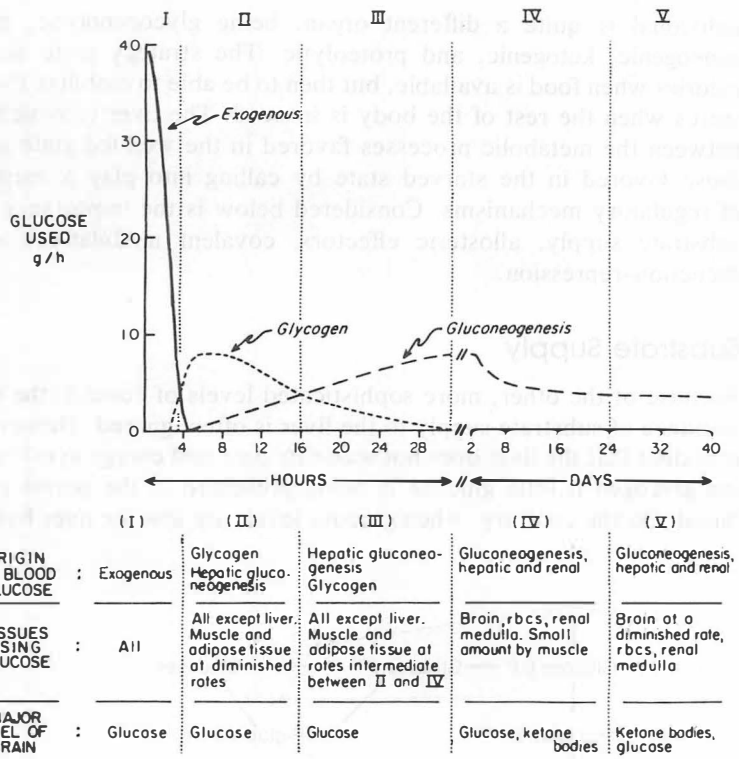


Figure 14.5
The five phases of glucose homeostasis in human beings.
From Ruderman, N. B., Aoki, T. T., and Cahill, G. F., Jr. Gluconeogenesis and its disorders in man, in R. W. Hanson, and M. A. Mehlman (eds.), *Gluconeogenesis, its regulation in mammalian species.* New York: Wiley, 1976, p. 515. Reproduced with permission.

enzymes will occur. This will continue until practically all of the fat is gone as a consequence of starvation. After all of it is gone, the body has to resort to the utilization of muscle protein completely. Before it's gone—you're gone (see Clin. Corr. 14.6).

14.3 MECHANISMS INVOLVED IN SWITCHING THE METABOLISM OF THE LIVER BETWEEN THE WELL-FED STATE AND THE STARVED STATE

The liver of the well-fed person is actively engaged in processes that favor the synthesis of glycogen and fat; such a liver is glycogenic, glycolytic, lipogenic, and cholesterogenic. The liver of the fasting

CLIN. CORR. 14.6 STARVATION

Starvation, or an overall deficit in food intake, including both calories and protein, leads to the development of a syndrome known as nutritional marasmus. Marasmus is a word of Greek origin meaning "to waste." Although not restricted to any age group, it is most common in children under 1 year of age. In developing countries early weaning of infants from breast milk is a common cause of marasmus. This may result from pregnancies in rapid succession, the desire of the mother to return to work, or switching to artificial formulas. Although the latter usually provides a complete and nutritious diet for a growing infant, economically deprived parents tend to dilute it with water to make it last longer. This practice leads to insufficient intake of calories. Likewise, diarrhea and malabsorption can develop if safe water and sterile procedures are not used.

In contrast to kwashiorkor (see Clin. Corr. 14.2), subcutaneous fat, hepatomegaly, and fatty liver are absent in marasmus. Insufficient calories are available for the synthesis and storage of fat. Diarrhea, muscle wasting, and infections are common. The consequences are the same as kwashiorkor, that is, poor physical, intellectual, and psychological development.

individual is quite a different organ, being glycogenolytic, gluconeogenic, ketogenic, and proteolytic. The strategy is to store calories when food is available, but then to be able to mobilize these stores when the rest of the body is in need. The liver is switched between the metabolic processes favored in the well-fed state and those favored in the starved state by calling into play a variety of regulatory mechanisms. Considered below is the importance of substrate supply, allosteric effectors, covalent modulation, and induction–repression.

Substrate Supply

Because of the other, more sophisticated levels of control, the importance of substrate supply to the liver is often ignored. However, it is clear that the liver does not waste its time and energy synthesizing glycogen if little glucose is being presented in the portal vein blood. On the contrary, when glucose levels are low the liver has to

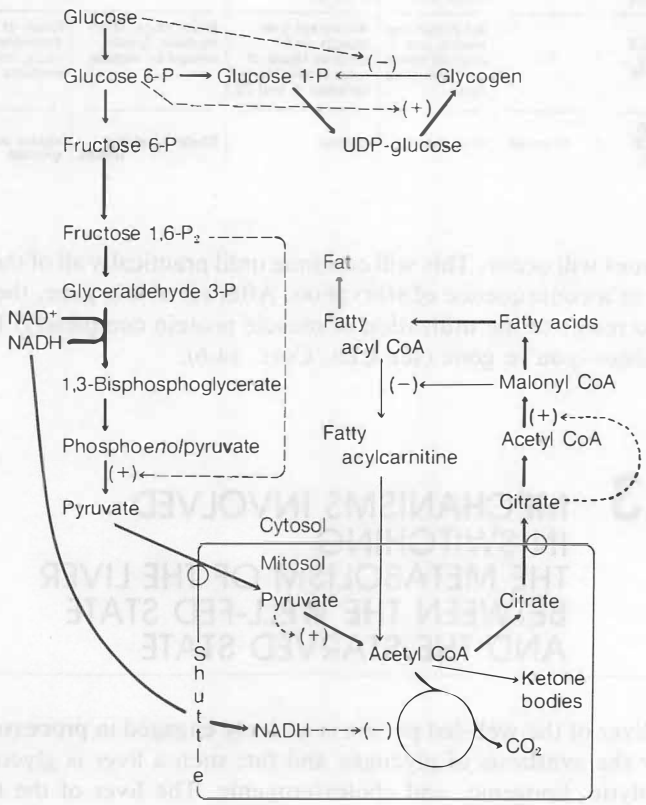


Figure 14.6
Control of hepatic metabolism in the well-fed state by allosteric effectors.

sense that it should mobilize its glycogen stores and synthesize glucose de novo in order to maintain adequate blood levels of glucose for the rest of the body. Likewise, the concentration of fatty acids in the blood of the portal vein is a major determinant of the rate of ketogenesis. As yet another example, excess fat is not synthesized and stored unless a person consumes excessive amounts of substrates that can be used for the process of lipogenesis.

It should also be pointed out that glucose synthesis by the liver is restricted by the rate at which gluconeogenic substrates (e.g., lactate, pyruvate, alanine, and glycerol) flow to the liver. In fact, the delivery of substrate to the liver represents an important mechanism for the maintenance of proper glucose levels. Delivery of excess amino acids to the liver of the diabetic, because of the accelerated and uncontrolled proteolysis (muscle wasting) that occurs in this metabolic condition, increases the rate of gluconeogenesis and exacerbates the hyperglycemia characteristic of diabetes. On the other hand, failure to supply the liver adequately with glucogenic substrate explains some types of hypoglycemia, such as that observed during pregnancy or advanced starvation.

From the examples provided above we can readily conclude that substrate supply is a major determinant of the rate at which virtually every metabolic process of the body operates. However, variations in substrate supply are clearly not sufficient to account for the tremendous changes in the direction of metabolism that must occur in the starve–feed cycle. As discussed below, regulation by allosteric effectors plays an important role in the different states.

Allosteric Effectors

Figures 14.6 and 14.7 summarize the effects of some of the negative and positive allosteric effectors believed to be important in the fed and starved states, respectively. As shown in Figure 14.6, glycogen synthesis is favored by the negative effects of glucose on glycogen phosphorylase; glycolysis is favored by fructose 1,6-bisphosphate activation of pyruvate kinase; pyruvate dehydrogenase activity is favored by pyruvate inhibition of the protein kinase, which phosphorylates and inactivates the pyruvate dehydrogenase complex; lipogenesis is favored by citrate activation of acetyl CoA carboxylase; and fatty acid oxidation is prevented by malonyl CoA inhibition of carnitine acyltransferase I activity. As shown in Figure 14.7, gluconeogenesis is favored in the starved state by acetyl CoA activation of pyruvate carboxylase and inhibition of pyruvate dehydrogenase, glycogenolysis is favored because of lower glucose levels and less inhibition of glycogen phosphorylase; lipogenesis is depressed because of inhibition of acetyl CoA carboxylase by long-

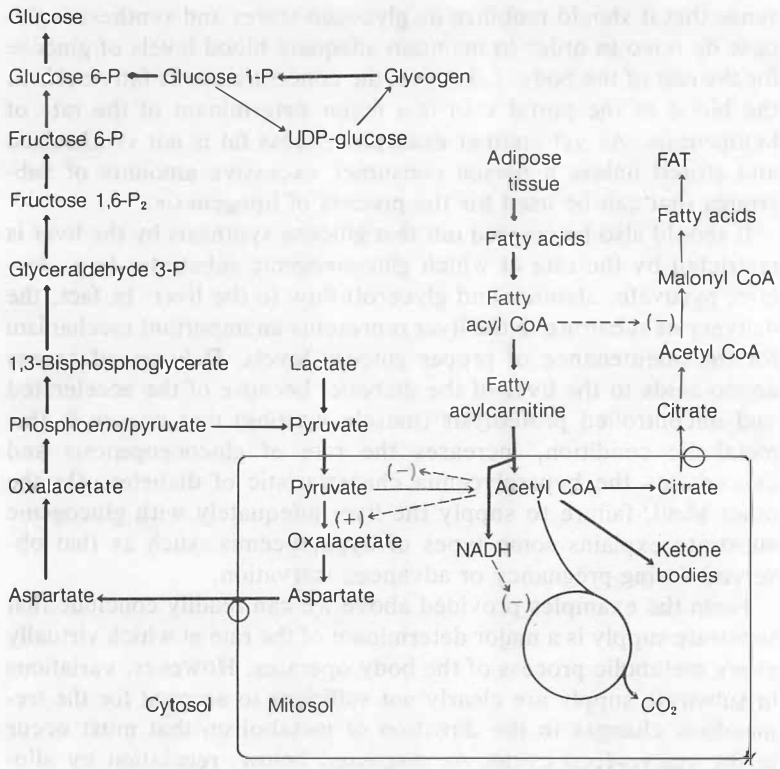


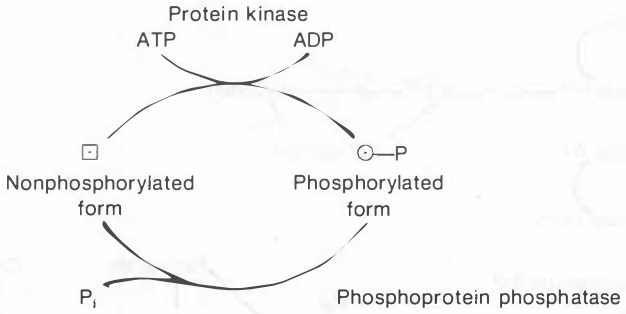
Figure 14.7
Control of hepatic metabolism in the fasting state by allosteric effectors.

chain acyl CoA esters; and fatty acid oxidation is accelerated because less malonyl CoA is present to inhibit carnitine acyltransferase I activity.

Allosteric effectors are important in switching the metabolic processes between the fed and starved states, but are thought to represent a more primitive control mechanism with respect to the degree of control that can be achieved by covalent modulation, summarized next.

Covalent Modulation

Figures 14.8 and 14.9 point out the interconvertible enzymes which are believed to play important roles in switching the liver between the well-fed and starved states, respectively. The regulation of enzymes by covalent modulation has been discussed in Chapter 7. Recall the following diagram, in which ⊡ and ⊙—P represent interconvertible forms of an enzyme in the nonphosphorylated and phosphorylated states, respectively.



The important points are (1) enzymes subject to covalent modulation undergo phosphorylation on one or more seryl residues by a protein kinase; (2) the phosphorylated enzyme can be returned to the dephosphorylated state by the action of a phosphoprotein phosphatase; (3) phosphorylation of the enzyme changes its conformation and its catalytic activity; (4) some enzymes are active only in the dephosphorylated state, others only in the phosphorylated state; (5) cAMP is the messenger that signals the phosphorylation of many, but not all, of the enzymes subject to covalent modulation; (6) cAMP acts by activating protein kinase; (7) glucagon increases cAMP levels in the liver by activating adenylate cyclase; and (8) insulin, by an unknown mechanism, which probably involves formation of an intracellular messenger, opposes the action of glucagon and cAMP, and thereby promotes dephosphorylation of the interconvertible enzymes.

As shown in Figure 14.8, all eight of the hepatic enzymes currently believed subject to covalent modulation are in the dephosphorylated mode in the liver of the well-fed animal. This makes sense, insulin : glucagon ratios being high in the blood and the cAMP levels being low in the liver in this situation. This results in a low activity for the cAMP-dependent protein kinase and a low degree of phosphorylation of those enzymes subject to phosphorylation by this enzyme. It is important to realize, however, that not all eight interconvertible enzymes are subject to phosphorylation by a cAMP-dependent protein kinase. At the time of this writing it is clear that glycogen synthase, glycogen phosphorylase (via phosphorylase kinase), and pyruvate kinase are subject to regulation by the cAMP-dependent protein kinase. Whether there is a link to cAMP-dependent protein kinase for the other enzymes is a question of current research interest. Only one of the interconvertible enzymes—glycogen phosphorylase—is inactive in the dephosphorylated mode. All of the other identified interconvertible enzymes (glycogen synthase, phosphofructokinase, pyruvate kinase, pyruvate dehydrogenase, acetyl CoA carboxylase, β-hydroxy-β-methylglutaryl CoA reductase, and glycerol phosphate

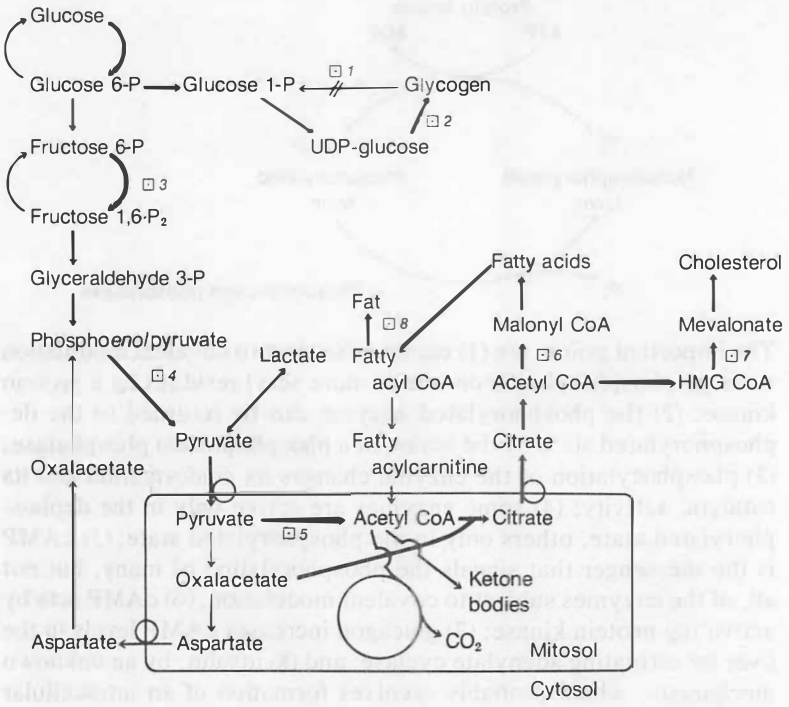


Figure 14.8
Activity and state of phosphorylation of the enzymes subject to covalent modulation in the lipogenic liver.
The dephosphorylated mode is indicated by the symbol ⊡. The interconvertible enzymes numbered are 1, *glycogen phosphorylase;* 2, *glycogen synthase;* 3, *phosphofructokinase;* 4, *pyruvate kinase,* 5, *pyruvate dehydrogenase;* 6, *acetyl CoA carboxylase;* 7, *β-hydroxy-β-methylglutaryl CoA reductase;* 8, *glycerol phosphate acyltransferase.*

acyltransferase) are active. Glycogenesis, glycolysis, and lipogenesis are greatly favored as a result of placing all of the interconvertible enzymes in the dephosphorylated mode. On the other hand, conditions are not right for the opposing pathways—glycogenolysis, gluconeogenesis, and ketogenesis.

As shown in Figure 14.9, all eight of the hepatic enzymes believed subject to covalent modulation are in the phosphorylated mode in the liver of the fasting animal. This also makes sense, insulin : glucagon ratios being low in the blood, cAMP levels being high, and the insulin "messenger" presumably being low in the liver in this situation. This results in greater protein kinase activity, and a greater extent of phosphorylation of those enzymes subject to phosphorylation by this kinase. In this case only one of the interconvertible enzymes—glycogen phosphorylase—is in its active catalytic state. All the other interconvertible enzymes are inactive in the phosphorylated mode. The result is that glycogenesis, glycolysis, and lipogenesis are shut down almost completely, and glycogenolysis, gluconeogenesis, and ketogenesis predominate.

Largely because it also contains enzymes subject to covalent modulation, adipose tissue responds just as dramatically as liver to

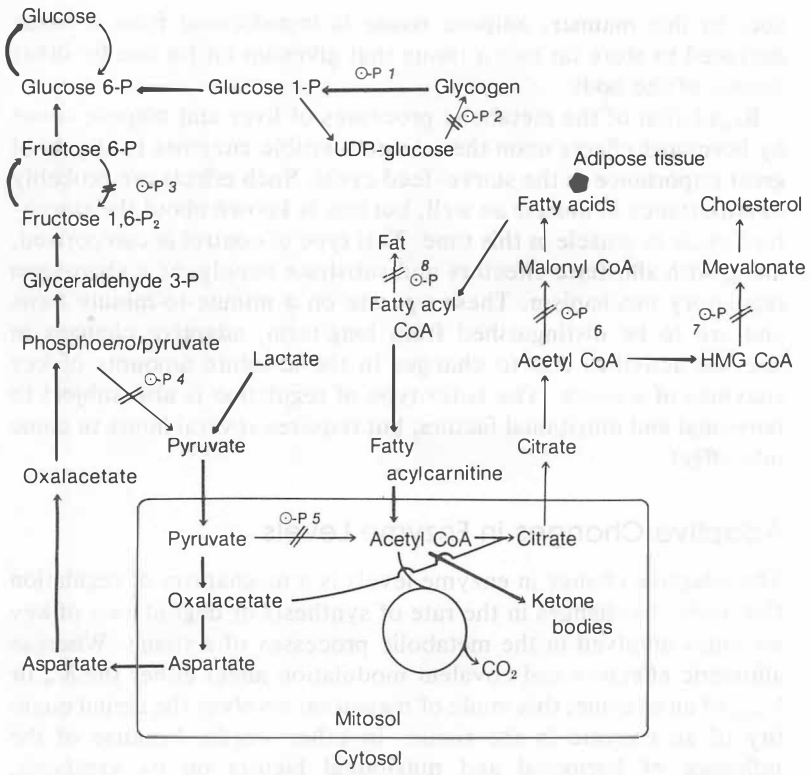


Figure 14.9
Activity and state of phosphorylation of the enzymes subject to covalent modulation in the glucogenic liver.
The phosphorylated mode is indicated by the symbol ⊙—P. The numbers refer to the same enzymes as in Figure 14.8.

the starve–feed cycle. Pyruvate kinase, pyruvate dehydrogenase, acetyl CoA carboxylase, and hormone-sensitive lipase (not found in liver) are all in the dephosphorylated mode in the adipose tissue of the well-fed person. As in liver, the first three enzymes are active in the dephosphorylated mode. Hormone-sensitive lipase is like glycogen phosphorylase, that is, inactive in the dephosphorylated mode. The high insulin : glucagon ratio, the low tissue cAMP concentration, and perhaps a high insulin "messenger" level are believed to be important determinants of the phosphorylation state of the interconvertible enzymes of this tissue. The result is that lipogenesis within adipose tissue is favored in the well-fed state. Like the liver, adipose tissue changes its metabolic characteristics in the starving individual. It lacks the enzyme machinery to synthesize much glycogen, or to participate in gluconeogenesis, but quickly shuts down lipogenesis and activates lipolysis in response to fasting. This is accomplished in large part by the phosphorylation of the enzymes subject to covalent modulation in adipose tissue. This is a consequence of the decrease in the insulin : glucagon ratio induced by fasting and the resulting increase in cAMP levels and protein kinase activity within the tis-

sue. In this manner, adipose tissue is transformed from a tissue designed to store fat into a tissue that gives up fat for use by other tissues of the body.

Regulation of the metabolic processes of liver and adipose tissue by hormonal effects upon the interconvertible enzymes is clearly of great importance in the starve–feed cycle. Such effects are probably of importance in muscle as well, but less is known about the starve–feed cycle in muscle at this time. This type of control is categorized, along with allosteric effectors and substrate supply, as a short-term regulatory mechanism. These operate on a minute-to-minute basis and are to be distinguished from long-term, adaptive changes in enzyme activities due to changes in the absolute amounts of key enzymes of a tissue. The latter type of regulation is also subject to hormonal and nutritional factors, but requires several hours to come into effect.

Adaptive Changes in Enzyme Levels

The adaptive change in enzyme levels is a mechanism of regulation that refers to changes in the rate of synthesis or degradation of key enzymes involved in the metabolic processes of a tissue. Whereas allosteric effectors and covalent modulation affect either the K_m or V_{max} of an enzyme, this mode of regulation involves the actual quantity of an enzyme in the tissue. In other words, because of the influence of hormonal and nutritional factors on its synthesis, there is physically more or less of a particular enzyme present in the tissue. For example, when a person is continuously maintained in a well-fed or overfed condition, the liver continuously improves in its capacity to synthesize fat. To be sure this can be explained in part by increased substrate supply, as well as appropriate changes in allosteric effectors (Figure 14.6) and the conversion of the interconvertible enzymes into the dephosphorylated mode (Figure 14.8). This is not the entire story, however, because the liver also has more molecules of those enzymes that play a key role in fat synthesis (see Figure 14.10). A whole battery of enzymes is induced, including glucokinase and pyruvate kinase for faster rates of glycolysis; glucose 6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, and malic enzyme to provide greater quantities of NADPH for reductive synthesis; and citrate cleavage enzyme, acetyl CoA carboxylase, fatty acid synthase, and Δ^9-desaturase for more rapid rates of fatty acid synthesis. All of these enzymes are induced to higher levels in the well-fed state, in response apparently to the increased insulin : glucagon ratios. While these enzymes are being induced, the opposite happens to those key enzymes that favor glucose synthesis. Phospho*enol*pyruvate carboxykinase, glucose

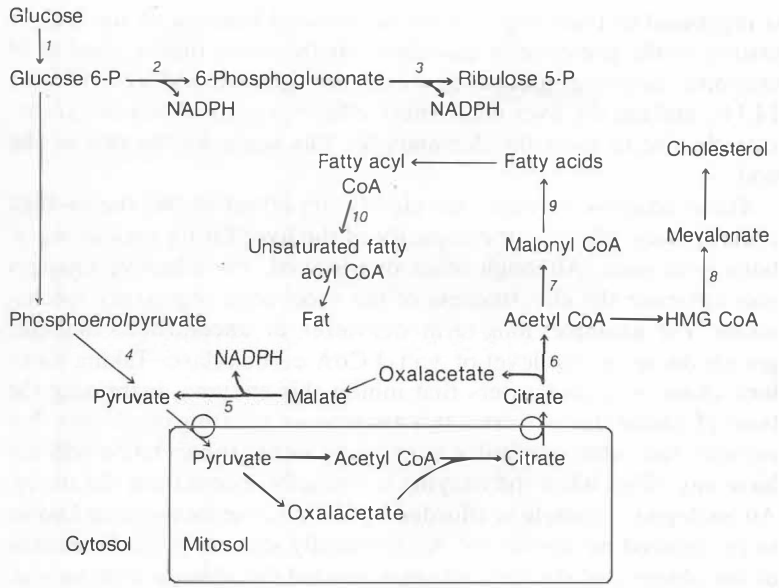


Figure 14.10
The battery of enzymes induced in the liver of the well-fed individual.
The inducible enzymes are numbered: 1, *glucokinase;* 2, *glucose 6-phosphate dehydrogenase;* 3, *6-phosphogluconate dehydrogenase;* 4, *pyruvate kinase;* 5, *malic enzyme;* 6, *citrate cleavage enzyme;* 7, *acetyl CoA carboxylase;* 8, *β-hydroxy-β-methylglutaryl CoA reductase;* 9, *fatty acid synthase; and* 10, Δ^9*-desaturase.*

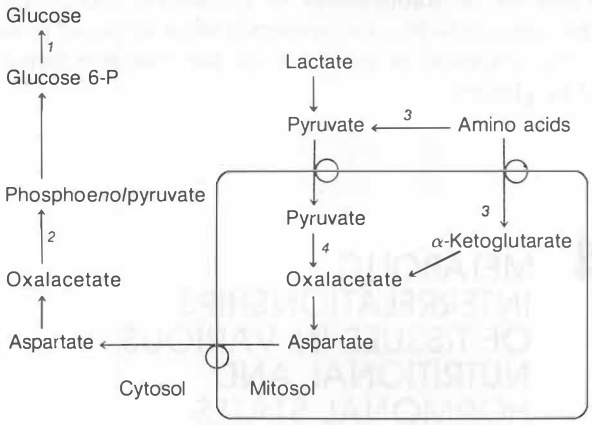


Figure 14.11
The battery of enzymes induced in the liver of an individual during fasting.
The inducible enzymes are numbered: 1, *glucose 6-phosphatase;* 2, *phosphoenolpyruvate carboxykinase;* 3, *various aminotransferases; and* 4, *pyruvate carboxylase.*

6-phosphatase, pyruvate carboxylase and some aminotransferases are decreased in amount, that is, their synthesis appears to be repressed in response to the presence of excess circulating glucose and insulin.

If a person fasts for several hours, the enzyme pattern characteristic of the liver changes dramatically (Figure 14.11). The enzymes favoring lipogenesis decrease in quantity, as though their synthesis

is repressed or their degradation accelerated because of the lack of insulin or the presence of glucagon. At the same time a number of enzymes favoring gluconeogenesis are greatly induced (Figure 14.11), making the liver much more effective in its ability to synthesize glucose to meet the demands for this sugar by the rest of the body.

These adaptive changes are clearly important in the starve–feed cycle, greatly affecting the capacity of the liver for its various metabolic processes. Although often overlooked, the adaptive changes also influence the effectiveness of the short-term regulatory mechanisms. For example, long-term starvation or uncontrolled diabetes greatly decreases the level of acetyl CoA carboxylase. Taking away long-chain acyl CoA esters that inhibit this enzyme, increasing the level of citrate that activates this enzyme, or creating conditions that activate this interconvertible enzyme by dephosphorylation will not have any effect when the enzyme is virtually absent from the tissue. An analogous example is afforded by the glucose intolerance known to be induced by starvation. A chronically starved person, because of the absence of the key enzymes needed for glucose metabolism, cannot effectively utilize a sudden load of blood glucose. A glucose load will set into motion, however, the induction of the required enzymes and the reestablishment of short-term regulatory mechanisms. This appears to be a consequence of an action of insulin upon the liver, the secretion of insulin from the pancreas being greatly increased by glucose.

14.4 METABOLIC INTERRELATIONSHIPS OF TISSUES IN VARIOUS NUTRITIONAL AND HORMONAL STATES

Many of the changes that occur in various nutritional and hormonal states of the human body are just variations on the starve–feed cycle, and completely predictable from what we have learned about the cycle. Some examples are given in Figure 14.12. Others are so obvious that a diagram is not necessary, for example, in rapid growth of a child, amino acids are in large part directed away from catabolism and into protein synthesis. On the other hand, the changes that occur in some physiologically important situations are

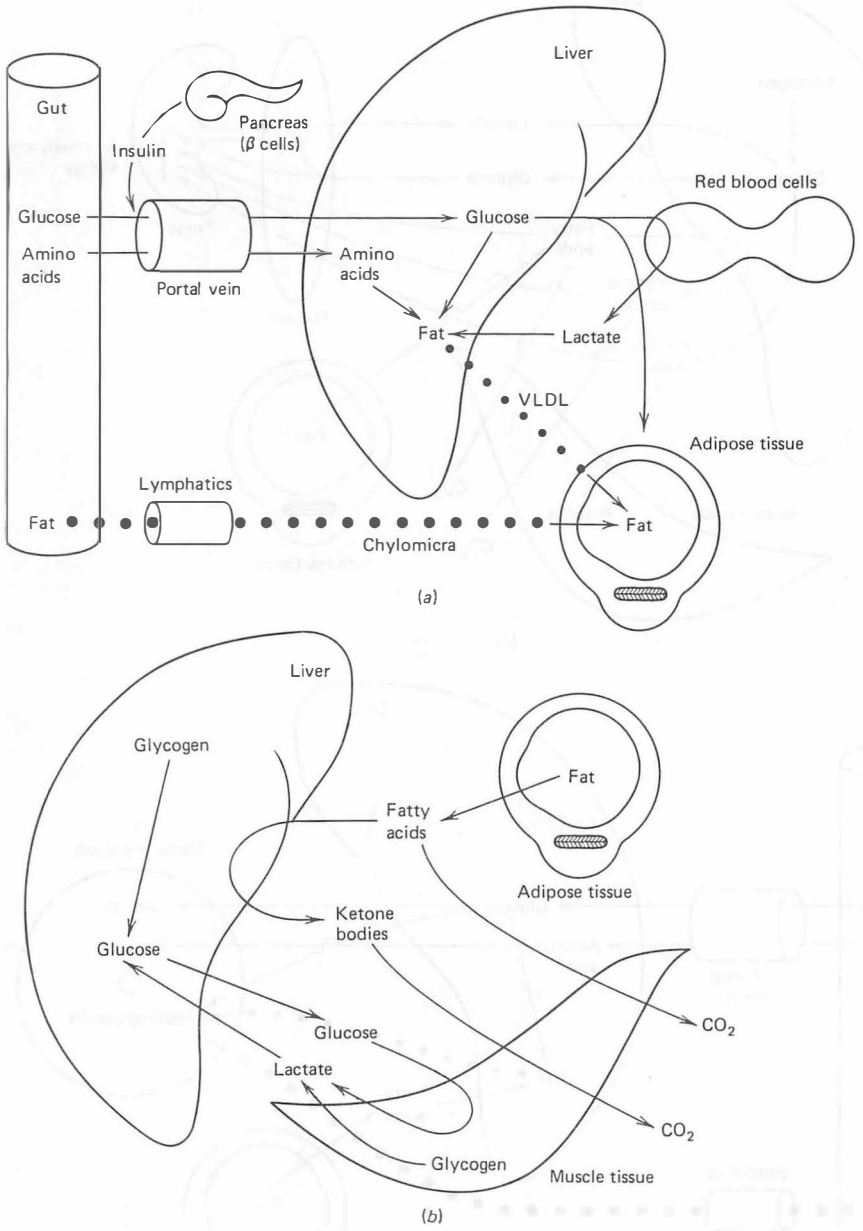
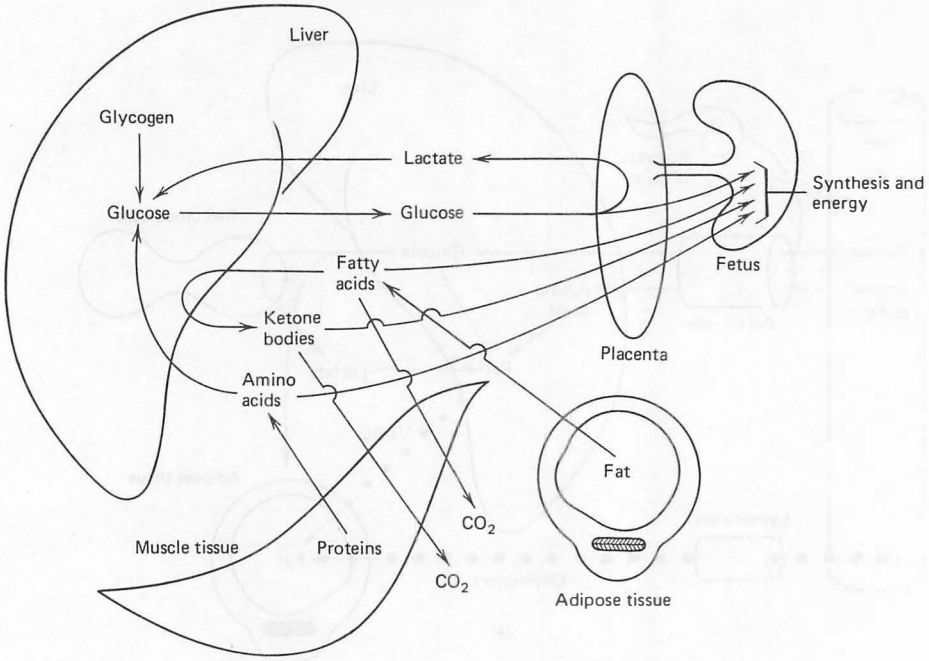
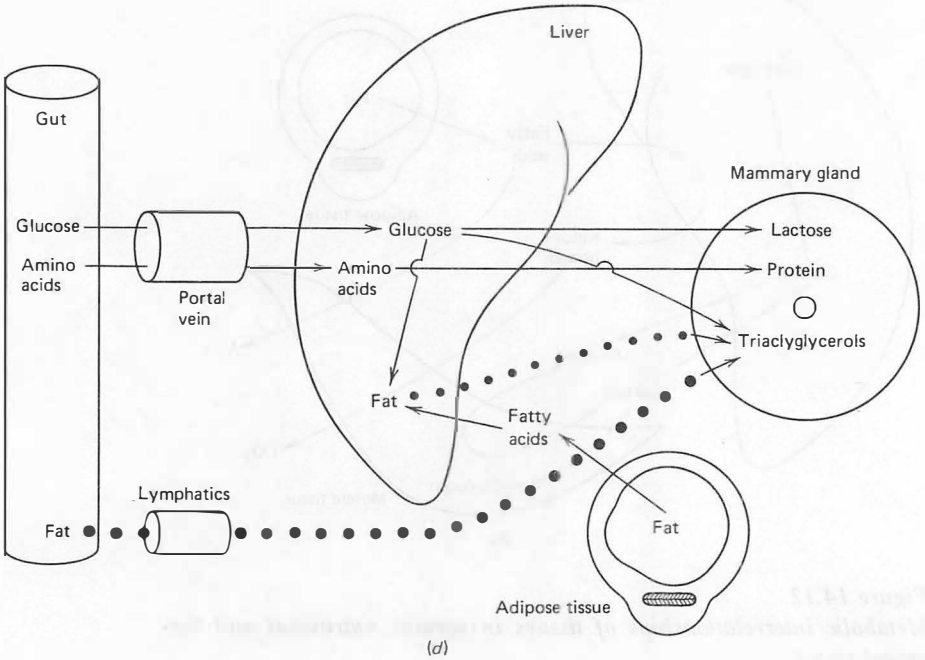


Figure 14.12
Metabolic interrelationships of tissues in various nutritional and hormonal states.
(a) Obesity. (b) Exercise. (c) Pregnancy. (d) Lactation. (e) Insulin-dependent diabetes mellitus. (f) Non-insulin-dependent diabetes mellitus (g) Stress. *Continued on pages 704–706.*

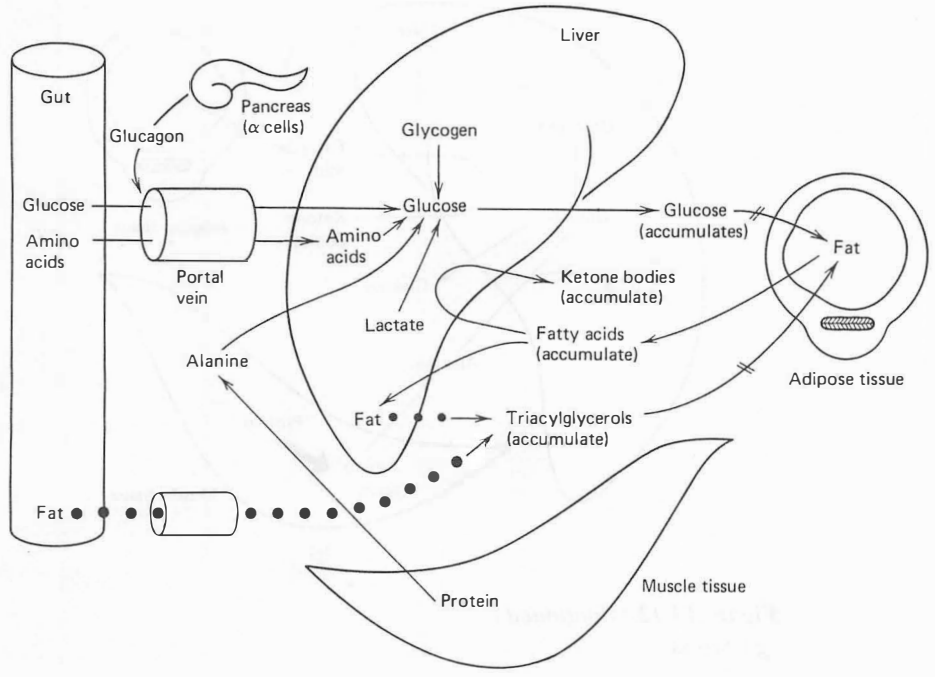


(c)

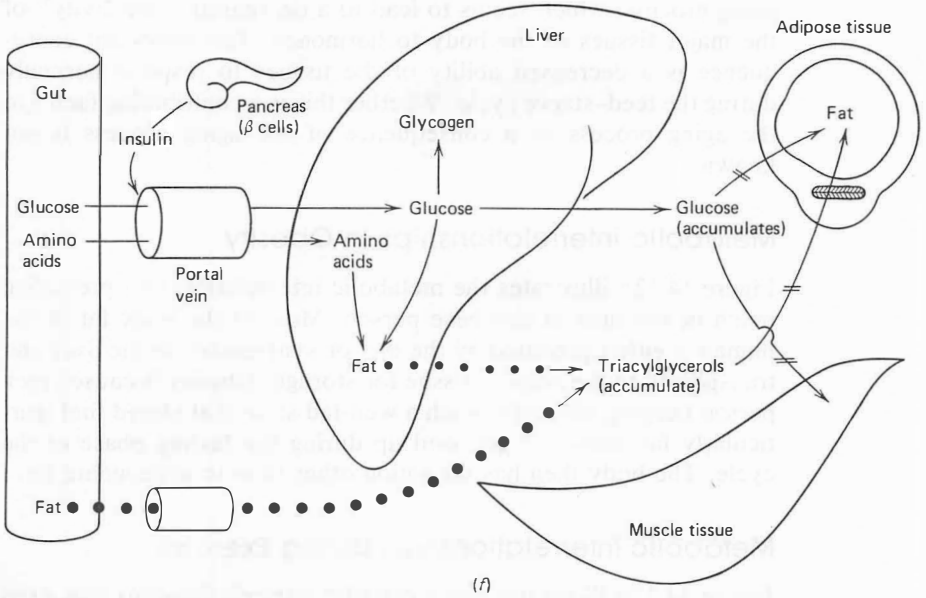


(d)

Figure 14.12 *(Continued)*
(c) Pregnancy. (d) Lactation.



(*e*)



(*f*)

Figure 14.12 *(Continued)*
(e) Insulin-dependent diabetes mellitus. (f) Non-insulin-dependent diabetes mellitus.

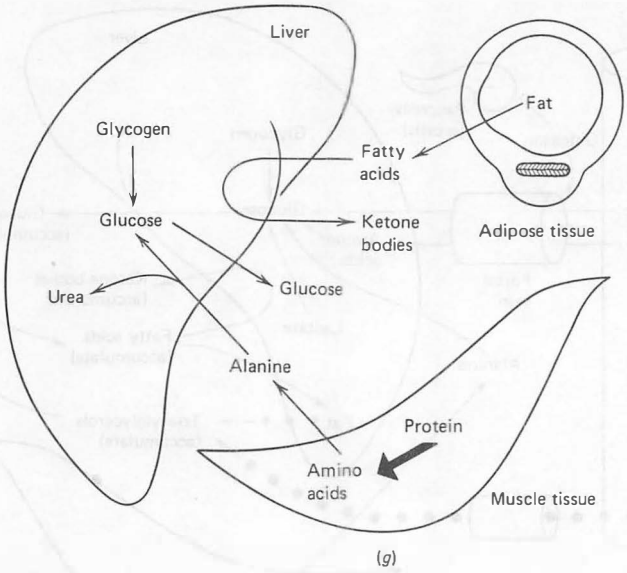


Figure 14.12 *(Continued)*
(g) Stress.

rather subtle and poorly understood. An example of the latter is the aging process, which seems to lead to a decreased "sensitivity" of the major tissues of the body to hormones. The important consequence is a decreased ability of the tissues to respond normally during the feed–starve cycle. Whether this is a contributing factor to the aging process or a consequence of the aging process is not known.

Metabolic Interrelationships in Obesity

Figure 14.12*a* illustrates the metabolic interrelationships prevailing much of the time in an obese person. Most of the body fat of the human is either provided by the diet or synthesized in the liver and transported to the adipose tissue for storage. Obesity is caused by a person keeping himself in such a well-fed state that stored fuel (particularly fat) does not get used up during the fasting phase of the cycle. The body then has no option other than to accumulate fat.

Metabolic Interrelationships During Exercise

Figure 14.12*b* illustrates the metabolic interrelationships that exist in heavy exercise. In some respects the changes are similar to those characteristic of the starved situation. In heavy exercise the liver

and adipose tissue adapt to supply the fuel needed for the work being performed by the muscles. The adipose tissue becomes lipolytic under the influence of epinephrine and glucagon in order to put more substrate into the blood in the form of fatty acids. The liver likewise becomes glycogenolytic, gluconeogenic, and ketogenic in order to supply the muscles with more glucose and ketone bodies. In turn, muscle utilizes these fuels along with its own glycogen reserves for the synthesis of ATP. As in the starved state, these changes in the metabolism of adipose tissue and liver are a consequence of short-term regulatory mechanisms in response to a decrease in the insulin : glucagon ratio. A major difference from the starved state is that muscle tissue uses both its glycogen reserves and blood glucose as substrates during exercise. The latter occurs in spite of the fact that low levels of insulin normally prevail during heavy exercise. Not understood at this time is the reason less insulin is needed for glucose uptake by exercising muscle. Of some clinical importance is the fact that exercise is beneficial in some insulin-requiring diabetics as a means of partially controlling blood glucose levels.

Metabolic Interrelationships During Pregnancy

The fetus represents another nutrient-requiring tissue (See Figure 14.12*c*). Glucose, ketone bodies, glycerol, amino acids, and probably fatty acids cross the placenta. Glucose and ketone bodies are major energy sources, but amino acid oxidation also occurs, as evidenced by placental release of ammonia and, near term, urea. Ammonia can also enter the fetal circulation, presumably to be used for amino acid biosynthesis. The placenta orchestrates maternal metabolism to provide for the fetus by the secretion of hormones. The sex steroids produce hyperinsulinemia and hyperlipidemia. Placental lactogen (or somatomammotropin) resembles growth hormone and stimulates lipolysis. In the postprandial period, especially during early pregnancy, blood glucose and liver glycogen are utilized more rapidly and the starved state is entered sooner. Referred to as "accelerated starvation," glucose and amino acid levels fall more rapidly than in the nonpregnant patient, and free fatty acids and ketone bodies increase, probably mediated by high rates of glucose and amino acid extraction by the fetus. The fetus plus placenta (conceptus) has some metabolic characteristics in common with muscle, in that it can use glucose and release lactate. The latter circulates back to the liver, establishing a Cori cycle analogous to that occurring between liver and muscle. Rapid switching between the fed and starved state is an even greater problem if the woman is also diabetic, because it is then very difficult to match insulin administration to the need for insulin, and hypoglycemia is common.

CLIN. CORR. 14.7 INSULIN-DEPENDENT DIABETES MELLITUS

Insulin-dependent diabetes mellitus is also called juvenile-onset diabetes because it usually appears in childhood or in the teens, but it is not limited to these patients. Insulin is either absent or nearly absent in this disease because of defective or absent β cells in the pancreas. Uncontrolled, it is characterized by hyperglycemia, hyperlipoproteinemia (chylomicra and VLDL), and episodes of severe ketoacidosis. Far from being a disease of defects only in carbohydrate metabolism, abnormalities exist in fat and protein metabolism in such patients as well. The hyperglycemia results in part from the inability of the insulin-dependent tissues to take up plasma glucose and in part by accelerated hepatic gluconeogenesis from amino acids derived from muscle protein. The ketoacidosis results from increased lipolysis in the adipose tissue and accelerated fatty acid oxidation in the liver. Hyperchylomicronemia is the result of low lipoprotein lipase activity in adipose tissue capillaries, an enzyme dependent upon insulin for its synthesis. For these reasons the disease is characterized by a "wasting away" of the flesh, leading ultimately to death if not treated soon after onset. Although insulin does not cure the diabetes, its use markedly alters the clinical course of the disease. The injected insulin promotes glucose uptake by tissues and inhibits gluconeogenesis, lipolysis, and proteolysis. The life span of the treated diabetic is still decreased by at least one-third, perhaps because it remains impossible to maintain perfect control of metabolism by repeated injections of insulin.

Later in pregnancy, the hormones released by the placenta oppose the effects of insulin, that is, promote lipolysis and hyperglycemia. The β cells respond with increased insulin secretion. The woman's metabolic state shifts then into that seen in obesity, and the rapid conversion between fed and starved modes is blunted.

Metabolic Interrelationships During Lactation

In late pregnancy placental hormones induce lipoprotein lipase in mammary gland and promote the development of milk-secreting cells and ducts. During lactation (see Figure 14.12*d*) the breast utilizes glucose for lactose and triacylglycerol synthesis. Amino acids are taken up for protein synthesis, and chylomicra and VLDL are utilized as sources of fatty acids for triacylglycerol synthesis. If these compounds are not supplied by the diet, proteolysis, gluconeogenesis, and lipolysis must supply them, resulting eventually in maternal malnutrition and poor milk quality.

Metabolic Interrelationships in Insulin-Dependent Diabetes Mellitus

Figure 14.12*e* shows the metabolic interrelationships that exist in juvenile-onset, insulin-dependent diabetes mellitus (see Clin. Corrs. 14.7 and 14.8). Because of defective β cell production of insulin, blood levels of insulin remain low in spite of elevated blood glucose levels. Even when dietary glucose is being delivered from the gut, the insulin : glucagon ratio cannot increase, and the liver remains gluconeogenic and ketogenic. Since it is impossible to switch to the processes of glycolysis, glycogenesis, and lipogenesis, the liver cannot properly buffer blood glucose levels. Indeed, since hepatic gluconeogenesis is elevated, the liver even contributes to the hyperglycemia in the well-fed state. The failure of many tissues to take up glucose in the absence of insulin contributes further to the hyperglycemia. Accelerated gluconeogenesis, fueled by substrate made available by body protein degradation, continues to maintain the hyperglycemia even in the starved state.

It may seem to be an enigma that hypertriglyceridemia is characteristic of this condition, since fatty acid synthesis is greatly diminished in the diabetic state. However, the low insulin : glucagon ratio results in uncontrolled rates of lipolysis in the adipose tissue. This increases blood levels of fatty acids and results in accelerated ketone body production by the liver. If the ketone bodies are not used as rapidly as they are formed, a dangerous condition develops, known as ketoacidosis, due to the excessive accumulation of ketone

bodies and protons. Regardless of whether ketoacidosis develops, not all of the fatty acid taken up by the liver can be handled by the pathway of fatty acid oxidation and ketogenesis. The excess is esterified and thereby directed into triacylglycerol and VLDL synthesis. Hypertriglyceridemia results because VLDL is being synthesized and released by the liver more rapidly than these lipid-laden particles can be cleared from the blood by lipoprotein lipase, which itself is dependent upon a high insulin : glucagon ratio. The defect in lipoprotein lipase also results in hyperchylomicronemia, since this enzyme is also required for chylomicron catabolism in adipose tissue.

The most important thing to remember about the diabetic state is that every tissue continues to play the catabolic roles that it was designed to play in starvation, in spite of delivery of adequate or even excess fuel from the gut. The consequence is that metabolism becomes stuck in the starve phase of the starve–feed cycle, with life-threatening consequences.

Metabolic Interrelationships in Non-Insulin-Dependent Diabetes Mellitus

Figure 14.12*f* shows the metabolic interrelationships characteristic of a person suffering from non-insulin-dependent, maturity-onset diabetes. In contrast to juvenile diabetes discussed above, insulin is not absent in maturity-onset diabetes (see Clin. Corr. 14.9). Indeed high levels of insulin may be observed in this form of diabetes, the problem being primarily one of insulin resistance of the tissues rather than lack of insulin. Insulin resistance simply means that insulin is present, but the normally responsive tissues—adipose tissue, muscle, liver, and heart—fail to respond, possibly because the insulin receptors of their plasma membranes are either decreased in number or affinity for insulin. The majority of patients with non-insulin-dependent diabetes mellitus are obese. Their insulin levels, although high, are not as high as a nondiabetic but similarly obese person (which also represents insulin resistance). Hence, this form of diabetes is also a form of β-cell failure, and exogenous insulin will reduce the hyperglycemia. Hyperglycemia results mainly because of poor uptake of glucose by peripheral tissues. In striking contrast to juvenile-onset diabetes, keotacidosis does not develop because uncontrolled lipolysis in the adipose tissue is not a feature of this disease. On the other hand, hypertriglyceridemia is characteristic of maturity-onset diabetes, but usually corresponds to an increase in VLDL without hyperchylomicronemia. The reason has to be quite different, most likely being explained by rapid rates of de novo hepatic synthesis of fatty acids rather than increased delivery of fatty acids from the adipose tissue.

CLIN. CORR. 14.8 VACOR-INDUCED DIABETES

N-3-Pyridylmethyl-*N'*-*p*-nitrophenylurea, or Vacor, was introduced in the mid-1970s as a rat poison, which would kill even Warfarin-resistant rats but not harm other animal life. The LD_{50} of Vacor is very low for rats (15 mg/kg body weight) but quite high for nontarget species, including dogs, cats, chickens, and little children. Unfortunately, Vacor poisoning in man has occurred after accidental and suicidal ingestion of the compound. Survivors of Vacor ingestion develop ketoacidosis-prone diabetes mellitus and severe toxic neuropathies. The former is probably caused by the destruction of β cells in the pancreatic islets of Langerhans, resulting in the development of insulin-dependent diabetes. Humans appear to be exquisitely sensitive to Vacor's diabetogenic effect, whereas Vacor-poisoned rats die of respiratory arrest and central nervous system dysfunction.

CLIN. CORR. 14.9 NON-INSULIN-DEPENDENT DIABETES MELLITUS

Non-insulin-dependent diabetes mellitus, which accounts for 80–90% of the diagnosed cases of diabetes, is also called maturity-onset diabetes to differentiate it from insulin-dependent, juvenile diabetes. It usually occurs in middle-aged obese people. Insulin is present at near-normal or even greatly elevated levels in this form of the disease. The defect or deficiency in these patients may be at the level of the insulin receptors located on the plasma membranes of normally insulin-responsive cells, that is, hepatocytes, adipocytes, and muscle cells. Non-insulin-dependent diabetes is characterized by hyperglycemia and hyperlipoproteinemia (VLDL). The ketoacidosis characteristic of the insulin-dependent disease is not observed. Increased levels of VLDL are probably the result of increased hepatic triacylglycerol synthesis stimulated by hyperglycemia and hyperinsulinemia. Obesity often comes before the development of insulin-independent diabetes and appears to be a major contributing factor. Overeating clearly causes the obesity, with the associated elevated insulin levels perhaps leading to the disease, for example, by an eventual repression of the synthesis of insulin receptors. An inverse relationship between insulin levels and the number of insulin receptors has been established. The higher the basal level of insulin, the fewer number of receptors present on the plasma membranes. The consequence is that insulin levels remain high, but glucose

Metabolic Interrelationships in Stress or Injury

Stress can include trauma due to injury, elective surgery, renal failure, burns, and infections. Characteristically there is an increase in blood cortisol, glucagon, catecholamines, and growth hormone. Although insulin secretion is not grossly impaired, insulin resistance is apparent. Foremost among the results of this is increased lipolysis and conversion of many tissues from glucose to fat oxidation (see Figure 14.12*g*). Ketogenesis is accelerated, which contributes to the change in energy metabolism. Although these changes are similar to those found in fasting, protein metabolism is grossly disordered. Instead of protein conservation, protein synthesis is decreased and proteolysis is greatly accelerated, with most of the amino acid carbon and nitrogen being sent to the liver for gluconeogenesis and urea synthesis. Thus negative nitrogen balance is characteristic of stressed patients. This may relate to the high levels of cortisol in stress conditions, which are absent in fasting. This condition is reversed when the patient recovers sufficiently from the illness to resume adequate caloric intake or if large amounts of carbohydrate, lipid, and amino acids are given by vein (total parenteral nutrition).

14.5 SUMMARY

This chapter has stressed the working relationship between liver, muscle, gut, and adipose tissue in the maintenance of caloric homeostasis during the starve–feed cycle. The liver functions at the center of this relationship, being able to switch from an organ primarily involved in the synthesis of glycogen and fat in the well-fed state into an organ primarily involved in the synthesis of glucose and ketone bodies in the fasting state. This dramatic transformation is brought about by the combined effects of several different regulatory mechanisms, including substrate supply, allosteric effectors, covalent modulation, and induction–repression. The metabolic response of the major tissues of the body to various diseases is often analogous to that characteristic of the starve–feed cycle.

BIBLIOGRAPHY

Cahill, G. F., Jr. Diabetes mellitus: a brief overview. *Johns Hopkins Med. J.* 143:155, 1978.

Cerami, A., and Koenig, R. H. Hemoglobin A_{Ic} as a model for the development of the sequelae of diabetes mellitus. *Trends Biochem. Soc.* 3:73, 1978.

Cohen, P. *Control of enzyme activity.* New York: Wiley, 1976.

Denton, R. M., and Pogson, C. I. *Metabolic regulation.* New York: Wiley, 1976.

Geelen, M. J. H., Harris, R. A., Beynen, A. C., and McCune, S. A. Short-term hormonal control of hepatic lipogenesis. *Diabetes* 29:1006, 1980.

Gibson, D. M., Lyons, R. T., Scott, D. F., and Muto, Y. Synthesis and degradation of lipogenic enzymes of rat liver. *Adv. Enz. Reg.* 10:187, 1972.

Karam, J. H., Lewitt, P. A., Young, C. W., Nowlain, R. E., Frankel, B. J., Jugiya, H., Freedman, Z. R., and Grodsky, G. M. Insulinopenic diabetes after rodenticide (vacor) ingestion: a unique model of acquired diabetes in man. *Diabetes* 29:971, 1980.

Klachko, D. M., Anderson, R. R., and Heimberg, M. (eds.). *Hormones and energy metabolism.* New York: Plenum, 1979.

Krebs, H. A. The effects of ethanol on the metabolic activity of the liver. *Adv. Enz. Reg.* 6:467, 1968.

Krebs, H. A. Some aspects of the regulation of fuel supply in omnivorous animals. *Adv. Enz. Reg.* 10:387, 1972.

Krebs, H. A., Williamson, D. H., Bates, M. W., Page, M. A., and Hawkins, R. A. The role of ketone bodies in caloric homeostasis. *Adv. Enz. Reg.* 9:387, 1971.

Larner, J. *Intermediary metabolism and its regulation.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

McGarry, J. D. New perspectives in the regulation of ketogenesis. *Diabetes* 28:517, 1979.

Newsholme, E. A., and Start, C. *Regulation in metabolism.* New York: Wiley, 1973.

Pregnancy metabolism, diabetes and the fetus. Ciba Foundation Symposium 163. Amsterdam: Excerpta Medica, 1979.

Roth, J. Insulin receptors in diabetes. *Hosp. Pract.* May 1980, p. 98.

levels are poorly controlled because of the lack of normal responsiveness to insulin. Although the insulin level is high, it is not as high as in a person who is obese but not diabetic. In other words, there is a relative deficiency in the insulin supply from the β cells. Diet alone can most often control the disease in the obese diabetic. If such a person can be motivated to lose weight, insulin receptors will increase in number, which, in turn, will bring about an increase in both glucose tolerance and tissue sensitivity to insulin. The non-insulin-dependent diabetic has a less severe form of the disease but nevertheless develops the same complications as the insulin-dependent diabetic, that is, neuropathy, retinopathy, kidney disease, and coronary artery disease.

Ruderman, N. B., Aoki, T. T., and Cahill, G. F., Jr. Gluconeogenesis and its disorders in man, in R. W. Hanson, and M. A. Mehlman (eds.), *Gluconeogenesis, its regulation in mammalian species.* New York: Wiley, 1976, p. 515.

Williamson, D. H., and Whitelow, E. Physiological aspects of the regulation of ketogenesis. *Biochem. Soc. Symp.* 43:137, 1978.

15

Biochemistry of Hormones I

FRANK UNGAR

HORMONE RECEPTORS, STEROID AND THYROID HORMONES

15.1 BIOCHEMICAL REGULATION

Enzyme Synthesis and Activation

Regulation of biochemical events occurs at all levels of biological organization, both inter- and intracellularly. The controlled entry of required substrates and the exclusion of noxious agents are fundamental properties of the cell. These events are controlled by the permeability characteristics of the membrane and by specific transport systems within the cell. Nuclear activity and the energy-producing systems within the cell, which depend on the access of suitable substances for metabolism, in turn regulate the production of protein and lipid components of the membrane and supply the energy to maintain these same protective membrane barriers. At any instant in time there is an interdependence among nuclear, cellular, and membrane events, which involves a communication network of substances of varying forms and modes of action. Every action is in response to a preceding reaction, which sets the pattern for subsequent events. In most cases we are able to define these changes in ionic and substrate concentrations or in enzyme activities as a result of chemical or physical modifications. The biochemist's approach to investigating this complex of dynamic causal relationships is usually to assume as a first approximation that a steady-state condition for each reaction can be defined, that is, $dx/dt = 0$, and then to quantitate the change that occurs after perturbation of the steady state. Many of the changes in cellular activity are modified by extracellular signals or messages and involve chemical and physical modifications of simple substances and macromolecules within the cell. The purpose of this chapter and Chapter 16 is to classify by name and function those various extracellular agents called hormones that are specifically involved in the transmission of these messages. While reading these sections, it would be well for the reader to refer to Chapter 14 and to review the material on metabolic control.

In addition to altering the flow of substrate within the cells, and in the turnover of enzyme protein, cellular activity is modified at the enzyme level essentially by five basic mechanisms:

1. Chemical modification of enzyme protein by phosphorylation and dephosphorylation, as in phosphorylase and glycogen synthetase activations
2. Cleavage of a covalent bond in the polypeptide chain, as in the conversion of trypsinogen to trypsin, and proinsulin to insulin
3. Association by aggregation, or disassociation of peptide subunits,

as in removal of a regulatory subunit by cAMP to activate protein kinase

4. Allosteric modification, which leads to a conformational change, such as the effect of concentration levels of AMP and ATP on phosphofructokinase and fructose 1,6-bisphosphatase
5. Cofactors, small molecules, or ions acting on enzyme complexes as prosthetic groups or activators, as in Ca^{2+}- and calmodulin-regulated events

Examples of these mechanisms are given in the succeeding sections on the alteration of cellular responses to hormone stimulation. Hormones are best thought of as modifying agents; they do not directly initiate or terminate enzyme activity. In addition, it is worth noting that expression of enzyme activity, once initiated, is the same irrespective of the source of stimulation, endocrine or nonendocrine.

In response to intra- or extracellular environmental signals, the cell can alter its activities with respect to growth and differentiation by increasing or decreasing enzyme or protein synthesis. Induction of enzymes or proteins occur through mechanisms that involve the DNA-directed synthesis of mRNA as proposed in the Jacob–Monod model, described in detail in Chapter 20.

The flow of this information could be controlled effectively at each of the five steps involving initiation, transcription, translation, posttranslation and degradation, any one of which could be the possible site of hormone action (Figure 15.1). Considerable evidence places a site of hormonal control at the level of initiation and transcription. In the bacterial cell enzyme suppression by glucose, and enzyme induction by lactose in the absence of glucose, of the β-galactosidase system by the *lac* operon in *E. coli* is one model used

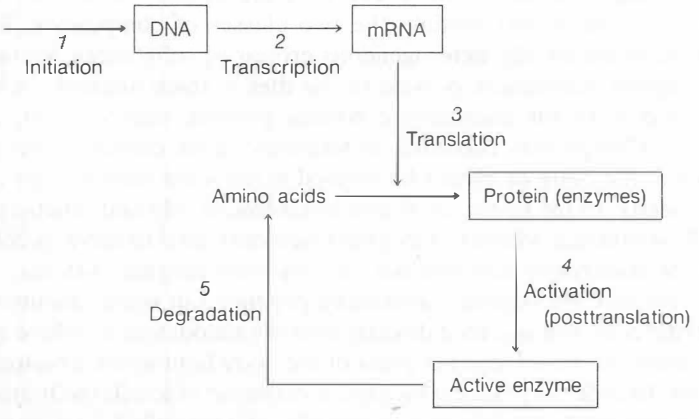


Figure 15.1
Sites of control of active enzyme levels.

for the study of transcriptional control. cAMP and a receptor protein for cAMP is involved in this regulation. In the mammalian system the formation of lactalbumin by the action of the steroid hormone, estradiol, in the uterine cell has provided the most complete information of control at the level of the genome.

In the human a number of the regulatory systems now have been defined, which aid communication within cells and between groups of cells or tissues. Whereas the microorganisms have limited defenses against changes in the environmental medium, the multicellular structure is so organized that the extracellular environment itself can be controlled, and communication links in this environment over great distances can be maintained. This control consists of both neural and hormonal elements.

Neuronal and hormonal regulation are best treated by an interdisciplinary approach, requiring a thorough understanding at the morphological level and in terms of molecular biology. A number of agents involved in these regulatory systems are described. To understand their involvement in biochemical regulation we first must characterize these compounds chemically, and then describe their role with respect to hormone action at the molecular level.

Definitions

The group of compounds discussed in this and the next chapter are involved in the control and regulation of cellular activity. These agents, which have been designated as *hormones,* are listed in Table 15.1 according to the gland of origin and primary function. The role of the hormones in metabolic activity and their interactions comprise a tightly controlled organized pattern, which in its entirety is commonly referred to as the *endocrine system.*

It is useful at this point to compare the actions of *vitamins* and *hormones* and to differentiate the two classes of compounds. Both consist of chemically heterogeneous groups of substances. *Vitamins* are organic substances present in the diet in trace amounts, which are necessary for maintaining normal growth, reproduction, and health. Compounds classified as vitamins either cannot be synthesized in the body or cannot be formed in amounts sufficient for normal needs. In the instances where biochemical role and function are well established, vitamins have been demonstrated to serve as cofactors or coenzymes and interact directly with enzyme systems.

Hormones are organic substances produced in trace amounts by specific cells and secreted directly into the bloodstream, where they circulate and travel to other parts of the body to produce a biological effect. In order to induce a biological response in a cell, the hormone must first bind as a ligand to a specific protein called a *receptor.*

Table 15.1 The Principal Hormones and Their Actions

Gland of Origin	*Hormone Name (Symbol)*	*Primary Actions*
Hypothalamus	Thyrotropin releasing factor (TRF)	Release of pituitary thyrotropin (TSH)
	Gonadotropin releasing factor (LH/FSH-RF)	Release of both pituitary FSH and LH
	Somatostatin or somatropin release inhibitory factor (SRIF)	Inhibits the release of pituitary growth hormone
(Stored in posterior pituitary)	Vasopressin (antidiuretic hormone, ADH)	Contraction of blood vessels, kidney reabsorption of water
	Oxytocin	Stimulates uterine contraction, milk ejection
	Vasotocin	Maintains water balance (nonmammalian species)
(Median eminence)	Melanocyte stimulating hormone (MSH)	Dispersion of pigment granules
Anterior Pituitary	Somatotropin or growth hormone (STH or GH)	Growth of body, organs, and bones
	Thyrotropin (TSH)	Size and function of thyroid
	Adrenocorticotropic hormone (ACTH)	Size and function of adrenal cortex
	Follicle stimulating hormone (FSH)	Growth of Graafian follicle, spermatogenesis (with LH)
	Luteinizing hormone (LH), interstitial cell stimulating hormone, (ICSH)	Causes ovulation *with* FSH, formation of testosterone and progesterone in interstitial cells
	Prolactin, mammotropin (luteotropin)	Growth of mammary gland, lactation, corpus luteum function
	Lipotropin (fat-mobilizing factor)	Release and oxidation of fats from adipose tissue
Parathyroid	Parathyroid hormone	Increases blood calcium Excretion of phosphate by kidney

Table 15.1 (*Continued*)

Gland of Origin	*Hormone Name (Symbol)*	*Primary Actions*
Parathyroid and thyroid	Calcitonin	Lowers blood calcium
Thyroid	Thyroxine (T_4) Triiodothyronine (T_3)	Growth and maturation, and metabolic rate Metamorphosis
Pancreatic islets		
β cells	Insulin	Hypoglycemic factor Regulation of CHO, fats, proteins
α cells	Glucagon	Liver glycogenolysis
Adrenal medulla	Epinephrine	Liver and muscle glycogenolysis
	Norepinephrine	
Adrenal cortex	Cortisol	Carbohydrate metabolism
	Aldosterone	Mineral metabolism
	Adrenal androgens	Androgenic activity (esp. females)
Pineal gland (epiphysis)	Indoles, serotonin, and melatonin	Effects on biological rhthyms and brain function Counteracts MSH activity
Ovaries	Estrogens	Estrus cycle, female sex properties
	Progesterone	Secretory phase (with estrogens) of uterus and mammary glands
	Relaxin	Relaxes symphysis pubis for birth
Testis	Testosterone and androgens	Male sex properties and spermatogenesis
Placenta	Placental lactogenic hormone	Growth hormone–prolactin activity
	Chorionic gonadotropin, estrogen, progesterone	Adjunct to other endocrine glands in 2nd and 3rd stages of pregnancy
Kidney	Renin	Hydrolysis of blood precursor protein to yield angiotensin

Table 15.1 (*Continued*)

Gland of Origin	*Hormone Name (Symbol)*	*Primary Actions*
Prostate, gonads, many tissues	Prostaglandins (PG)	Many effects at membrane site of synthesis
Gastrointestinal (GI) tract	Gastrin	Stimulates parietal cell secretions
	Secretin	Stimulates pancreatic juice
	Cholecystokinin	Contraction of gall-bladder
Brain opioids	Endorphins: β-endorphin, enkephalins	Endogenous, peptides which bind to morphine receptor

The classification of substances as hormones or vitamins has been established primarily by historical precedent and may be modified as new information is acquired. It should be recognized that thiamine, riboflavin, and other substances of the vitamin B group are classified as vitamins due to their dietary requirement in human nutrition, but in plants, where these substances can be synthesized, their actions on growth and development qualify these agents as true plant hormones. Ascorbic acid, or vitamin C, is a vitamin in human nutrition, and in the guinea pig as well, because an enzyme in the synthesis from gulonolactone to ascorbic acid is missing. Since the rat has this enzyme and ascorbic acid can be synthesized in its tissues, ascorbic acid is not a vitamin for this species. Vitamin D_3, or cholecalciferol, has been classified as a vitamin in human nutrition based on its low conversion from a precursor sterol under limiting (no sunlight) conditions. It is now recognized that the cholecalciferol precursor, which originates in cells in the skin, is modified by liver and kidney action to a more active form, which stimulates Ca^{2+} ion absorption in the mucosal cells of the gastrointestinal (GI) tract. By virtue of its conversion by hydroxylation in a specific tissue to form a derivative that affects other cells of the body, vitamin D_3 can be considered to be a hormone.

Vitamin D_3 shares one other important attribute with substances classified as hormones. As is the case with other steroid compounds and with thyroid hormones, vitamin D_3 binds to receptors within the cell nucleus. Response of the cell to hormone stimulation in general

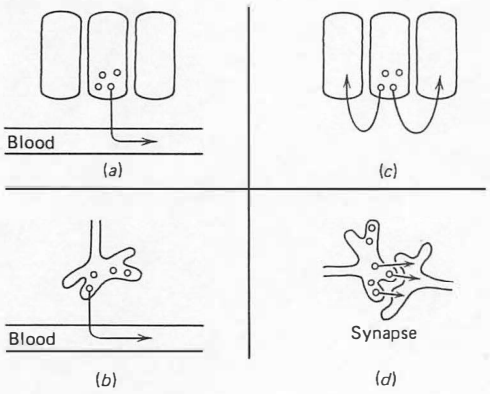


Figure 15.2
Secretory cells.
(*a*) *Endocrine cell.* (*b*) *Neuroendocrine cell.* (*c*) *Paracrine cell.* (*d*) *Nerve cell.*

is considered to involve gene activation with increased protein or enzyme synthesis due to the DNA-directed synthesis of messenger RNA. A large body of evidence has accumulated both for those hormones that bind to nuclear receptors and for the hormones that bind to cell plasma membrane receptors to support the concept of nuclear activation as an integral part of hormone action.

Secretory Cells

One further distinction may be useful in differentiating endocrine cells and other cell types and their secretions (Figure 15.2). Cells are referred to as *endocrine* cells (Figure 15.2*a*) if they produce and secrete active substances (hormones) directly into the bloodstream. The secretion of the hormone insulin by the pancreatic β cell is an example. *Neuroendocrine* cells (Figure 15.2*b*) are secretory nerve cells. The substance, for example, vasopressin, produced by nerve cells in the hypothalamus and secreted by the neurohypophysis directly into the bloodstream, is called a *neurohormone*. *Paracrine* cells (Figure 15.2*c*) produce substances that affect adjacent cells directly without transport into the circulating blood. Pancreatic somatostatin is an example. These substances by definition would not be considered hormones. Substances released by nerve cells (Figure 15.2*d*) into the synaptic cleft to stimulate contiguous nerve cells are called *neurotransmitters.* These substances, such as norepinephrine or acetylcholine, are secreted and inactivated locally at the synaptic cleft and therefore would not fit the classical definition of hormone. However, epinephrine and norepinephrine also are produced by the adrenal medulla and are secreted directly into the circulating blood. Since they have effects on muscle and liver carbohydrate metabolism, they traditionally have been listed as hormones.

Peptide Hormone Assays

When biologically active substances, particularly if they are peptides or proteins, are present in circulating plasma in trace amounts, the task of isolation and characterization is incomplete, and quantitation is difficult. The usual spectrophotometric or analytical procedures are not sensitive at the levels (10^{-7} to 10^{-11} M) at which most hormones circulate, so that radioactive tracers necessarily are employed. The isolation and assay of biologically active substances, if chemical structure is not known or when concentrations are extremely low, depend upon procedures which by necessity are indirect and less reliable. The assessment and validity of measurement of a trace substance therefore rests upon an accumulation of data from separate, independent analyses. The use of any one assay by

itself usually is not sufficient to provide precise information concerning peptides in tissues or in plasma. Therefore, whenever possible the combined use of three assay procedures—(1) bioassay, (2) radioimmunoassay, and (3) receptor assay—is used for more effective and reliable estimates of levels of biologically active peptides.

Bioassay

Bioassay is the most sensitive and is the definitive measure of a biologically active substance. The bioassay quantitates a specific biological response in an in vivo or in vitro system by measuring an amount of test substance compared with known amounts of authentic standards plotted on a log dose–response curve. Although the precision of the bioassay is inferior to the chemical/physical assay, the sensitivity and specificity can be very good when an appropriate biological response is chosen. Examples of well-known bioassay procedures include the lowering in vivo of blood glucose levels in the rat with varying doses of injected insulin; or measuring the contraction of cardiac muscle with varying doses of epinephrine added to a medium bathing the tissue in vitro.

Receptor Assays

Receptor assays consist of the displacement of a radioactive form of ligand by a nonlabeled ligand bound to the receptor protein of a target cell. The amount of radioactivity displaced under standardized conditions of time and temperature is proportional to the amount of added nonlabeled ligand. The displacement by nonlabeled insulin of [^{125}I]insulin bound to a fat cell ghost, or by nonlabeled ACTH of [^{125}I]ACTH bound to an adrenal cortical cell membrane are examples of receptor assays used to measure levels of hormone in the circulating plasma.

Radioimmunoassay

Radioimmunoassay has rapidly become the most frequently used assay due to its extreme level of sensitivity, specificity, and general versatility. It is also one of the least complicated assays to perform and as a result of these factors lends itself to automated procedures capable of measurement of large numbers of tests using small sample volumes. In analogy to the receptor assay, an ^{125}I-labeled hormone is used to compete with a nonlabeled hormone for binding to an antibody. Antibodies to protein hormones can be raised specifically when the hormone used as an antigen is injected into an animal. When the ligand of interest is not antigenic, that is, it cannot generate an antibody by itself, the substance, used as a hapten, is made antigenic by covalent linkage to a protein such as bovine serum albumin. The complex is then injected into an animal, such as a

rabbit or guinea pig, to form antibodies. The procedure is thus generally applicable for the assay of a variety of proteins, peptides, hormones, vitamins, drugs, and organic substances.

It should be apparent that the three assays applied to a given amount of unknown protein may yield different quantitative results. This is due to the fact that the determinants for each assay will depend upon different parts of the protein structure with different amino acids as binding or reacting sites. The measurement of protein will vary considerably from one assay procedure to another, depending on the specificity of the antibody or the receptor for the ligand, the presence of other competing ligands, or the loss or denaturation of part of the isolated protein ligand molecule.

15.2 THE RECEPTOR MODEL FOR HORMONE ACTION

Estradiol Binding in Uterine Tissue

The activity of a hormone at the target cell is determined by its concentration, by its binding affinity to a receptor, by the number of receptor sites occupied, and by the duration of binding. The successful demonstration of specific incorporation of a hormone by its target tissue was first achieved with the use of trace amounts of tritium-labeled estradiol of extremely high specific activity (Figure 15.3). A single injection in the rat permitted the accurate measurement of as little as 1 pg(10^{-12} g) of steroid in the tissues and blood. A high concentration of radioactivity was attained very rapidly in muscle, kidney, and liver, followed by a rapid decrease which paralleled that found in the blood. In contrast, the estrogen growth responsive tissues, uterus and vagina, achieved high levels of radioactivity, which were maintained for longer periods of time. The selective concentration of hormone by its target tissue demonstrated a unique trapping mechanism in the target cell.

This avidity of the target tissue for the hormone was due to a protein unique for that cell. The term *receptor* was applied to the protein, since it could be shown that (1) the binding was specific for estradiol, and (2) there was an associated stimulation of uterine growth in response to the estrogen binding. No chemical alteration of the steroid molecule occurred as a result of estradiol binding to its receptor. To fully characterize a cell receptor, specificity of binding to the receptor by the ligand and a biological response as a result of the ligand binding must be demonstrated. A substance that competes

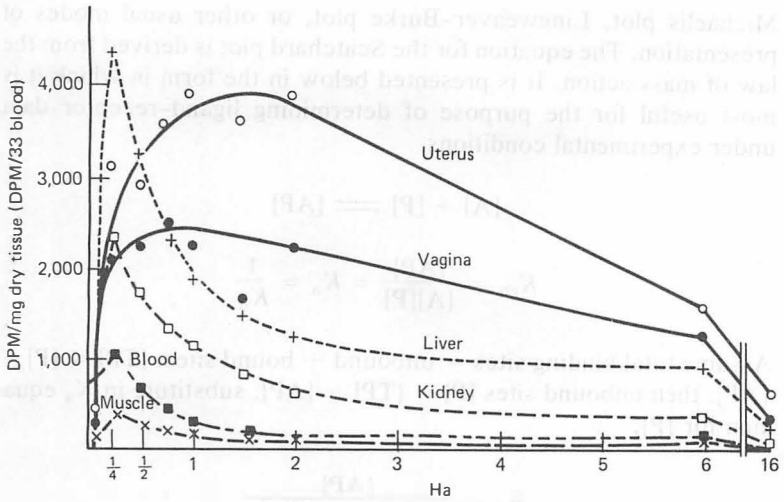


Figure 15.3
Tissue distribution of estradiol after a single subcutaneous injection of 6,7-^{3}H-estradiol.
Reproduced with permission from E. V. Jensen and H. I. Jacobson, *Recent Prog. Hormone Res.*, 18:387, 1962.

with a hormone for binding to a receptor is described as an *agonist* if the response of the cell is the same as, or mimics, the action of the hormone. A substance that competes with a hormone for binding to a receptor is referred to as an *antagonist* if it blocks or inhibits the response of the cell to the action of the hormone.

Due to the very low concentrations of specific receptor sites, tissues can easily be overwhelmed by excess ligand, in which case the high level of nonspecific binding would mask the specific receptor binding. A basic procedure to determine the binding characteristics of receptors, the number of binding sites, different classes of receptors, and the binding affinity or association constant of a ligand utilizes the Scatchard plot, described in the next section.

Scatchard Plot

The kinetics of ligand binding to a protein has been presented in the discussion of enzyme kinetics in Chapter 4. Except for the fact that there is no product involved, the assumptions and treatment of ligand–protein interaction are no different from that for the substrate–enzyme complex. Michaelis–Menten behavior is observed with receptor binding of ligand. Data can be arranged in a typical

Michaelis plot, Lineweaver–Burke plot, or other usual modes of presentation. The equation for the Scatchard plot is derived from the law of mass action. It is presented below in the form in which it is most useful for the purpose of determining ligand–receptor data under experimental conditions.

$$[A] + [P] \rightleftharpoons [AP]$$

$$K_{eq} = \frac{[AP]}{[A][P]} = K_a = \frac{1}{K_d}$$

Assume total binding sites = unbound + bound sites. [TP] = [P] + [AP]; then unbound sites [P] = [TP] − [AP]; substitute in K_a equation for [P],

$$K_a = \frac{[AP]}{[A]([TP] - [AP])}$$

Rearranging yields

$$\frac{\text{Bound}}{\text{Free}} = \frac{[AP]}{[A]} = K_a\,([TP] - [AP])$$

The equation is now set in the form of a straight line obtained by plotting [AP]/[A] vs [AP] (Figure 15.4*a*). Note that all of the parameters necessary for plotting values for the association constant (K_a) and for the total number of binding sites [TP] contains the ligand term (A). Using a radioactive form of ligand (^{3}H-A), it is then necessary only to separate the bound (protein-bound) ligand from the unbound (free) ligand by a suitable procedure and plot the radioactive counts along the ordinate and abscissa. Complete separation is

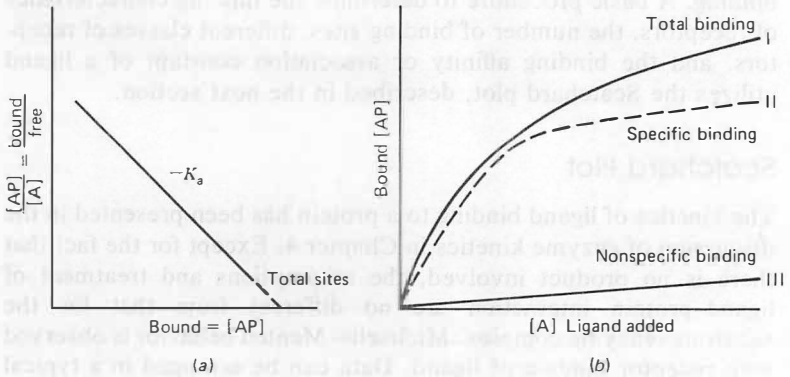


Figure 15.4
(a) Scatchard plot. Plotting the ratio of radioactivity in the protein-bound form (AP)/(A), the free ligand form, vs (AP) on the abscissa gives the slope = $-K_a$ (the association constant), and total sites (TP) by extrapolation to baseline. The points used to determine the straight line have been corrected for nonspecific binding by the procedure shown in (b). (b) Saturation analysis. Total radioactivity of ligand bound to protein (AP) is plotted vs radioactive ligand (A) added to give curve I. When a large excess of nonradioactive ligand is added with radioactive (A) to the same amount of receptor in the same volume, curve III is obtained. Curve II, specific binding is obtained by subtraction of values of curve III from curve I.

achieved by protein precipitation, adsorption of free form, chromatography, or electrophoresis. Since receptor concentrations and hormone (ligand) concentrations are very low (10^{-8} to 10^{-12} M), radioactive ligands must be used with exceedingly high (10–50 Ci/mM) specific activity.

The slope of the line (which is negative) gives the association constant ($-K_a$). Extrapolation to the abscissa gives the value for [TP] = total number of sites. In terms of Michaelis kinetics, the slope constant when half the sites are occupied, half unoccupied, is K_d, which is equivalent to K_m. The total number of ligand binding sites (maximum bound) is equivalent to V_{max}.

Receptor sites have the characteristics of being saturable, reversible, and specific for the ligand. Nonspecific binding of ligand has a relatively low affinity constant and nonsaturable kinetics. Therefore, specific binding is established by demonstrating high affinity and low capacity of receptor for the ligand. A correction for nonspecific binding can be made by diluting the radioactive ligand with a large excess of nonlabeled carrier ligand. Specific binding is equal to the total counts bound minus the correction factor for nonspecific binding (Figure 15.4*b*).

Steroid Hormone Receptors

When radioactive estradiol is administered to a female rat, most of the radioactivity subsequently is bound in the nucleus of the uterine cell. Estradiol first binds a cytoplasmic protein component upon entering the uterine cell. This steroid–receptor complex, after some modification of the protein, is then transferred to an acceptor site on the chromatin protein in the nucleus. The transfer from cytoplasm

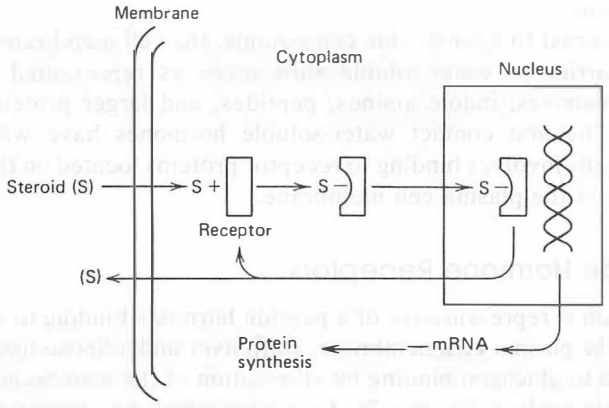


Figure 15.5
Steroid receptor model.

Steroid	*Target Cell Nucleus*
Estradiol	Uterus, breast, brain
Testosterone	Testes, brain
Cortisol	Liver, lymphocyte
Aldosterone	Kidney
Progesterone	Uterus, breast
Cholecalciferol	Intestinal mucosa

to nucleus, called translocation, involves both hormone and receptor being transported as a complex for nuclear binding. Each of the steroid hormones bind to nuclear sites in their respective target cells. Cytoplasmic receptor proteins with high affinity binding (K_a = 10^8–$10^{10}\,M^{-1}$) have been found for each steroid target cell (Figure 15.5).

Vitamin D and thyroid hormones also bind as a receptor complex to nuclear sites. The high affinity binding constants (K_a = $10^{11}\,M^{-1}$) for thyroid hormone in the liver cell nucleus indicates specific binding sites are present. However, no protein with high affinity for thyroid hormone has been found in the cell cytoplasm. Nuclear concentration of active forms of vitamin D have been demonstrated by radioautography in cells of bone, intestine, and kidney, presumed target cells for vitamin D.

The difference in sites of binding of steroid and thyroid hormones is represented as follows:

	Cytoplasm		*Nucleus*
Steroid hormone	⟶ steroid–receptor complex	⟶	steroid–receptor complex binding to nuclear acceptor

	Nucleus
Thyroid hormone ⟶	thyroid hormone binding to nuclear acceptor

Steroid hormones and thyroid hormones are lipid-soluble compounds and freely permeable to the plasma cell membrane. There appears to be no barrier to their entrance or exit from tissue cells. Uptake and subsequent response to steroid hormones are determined by the ratio of occupied to unoccupied receptor sites present in the cell. Upon release of steroid, receptor may be recycled in the cytoplasm.

In contrast to lipid-soluble compounds, the cell membrane serves as a barrier to water-soluble substances as represented by the catecholamines, indole amines, peptides, and larger protein molecules. The first contact water-soluble hormones have with their target cells involves binding to receptor proteins located on the outer surface of the plasma cell membrane.

Peptide Hormone Receptors

Glucagon is representative of a peptide hormone binding to a receptor on the plasma cell membrane. Both liver and adipose tissue cells respond to glucagon binding by stimulation of the membrane-bound adenylate cyclase (Chapter 7). An in vitro membrane preparation can

be used to study simultaneously the binding reaction using [^{125}I]-glucagon and the response as a result of that binding by measuring cAMP production. In addition, receptor specificity can be studied using the fat cell, since it is a target organ for glucagon, epinephrine, ACTH, lipotropin, and growth hormone. The response of the fat cell (release of free fatty acids) to a mixture of different hormones is additive, that is, there is an increased increment of release with the addition of each hormone, indicating separate receptor proteins for each hormone. In this system receptor binding results in the activation of adenylate cyclase as a common response for each stimulus. A number of hormone receptor sites on the outer surface of the membrane can stimulate a given adenylate cyclase enzyme situated on the inner (cytoplasmic) surface of the membrane (Figure 15.6). Further agents of receptor binding of individual peptide hormones are discussed in Chapter 16.

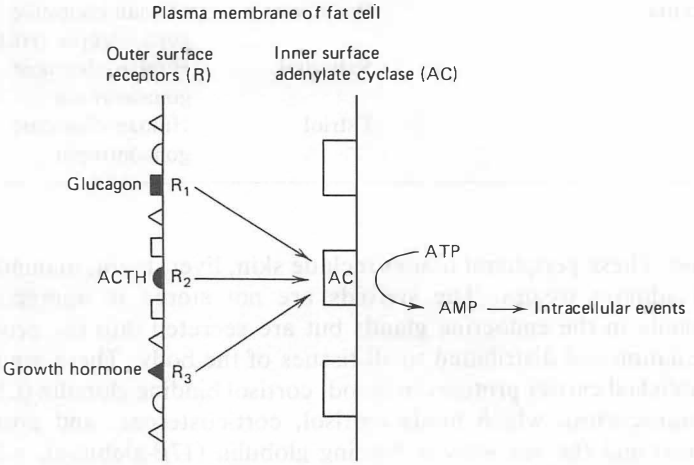


Figure 15.6
Receptor model of peptide hormones.

15.3 STEROID HORMONES

Metabolic Considerations

The steroid hormones are produced in specific cells of the adrenal cortex, the testis, the ovary, and placenta (Table 15.2). Biologically active steroids, particularly androgens and estrogens, are formed also in nonendocrine tissues from steroid precursors circulating in

Table 15.2 Tissues Source of Steroid Hormones

Organ	*Cell Type*	*Secretion*	*Control*
Adrenal cortex	Glomerulosa	Aldosterone	Angiotensin and Na/K ratio
	Fasciculata-reticularis	Cortisol Dehydroepian-drosterone (DHEA)	Adrenocorticotropic hormone (ACTH)
Testis	Leydig cell	Testosterone 4-Androstene-3,17-dione	Luteinizing hormone (LH)
Ovary	Follicle (theca)	Estradiol	Luteinizing hormone and follicle stimulating hormone
	Stroma	4-Androstene-3,17-dione	Luteinizing hormone
	Corpus luteum	Progesterone	Luteinizing hormone
Placenta		Progesterone	Human chorionic gonadotropin (HCG)
		Estradiol	Human chorionic gonadotropin
		Estriol	Human chorionic gonadotropin

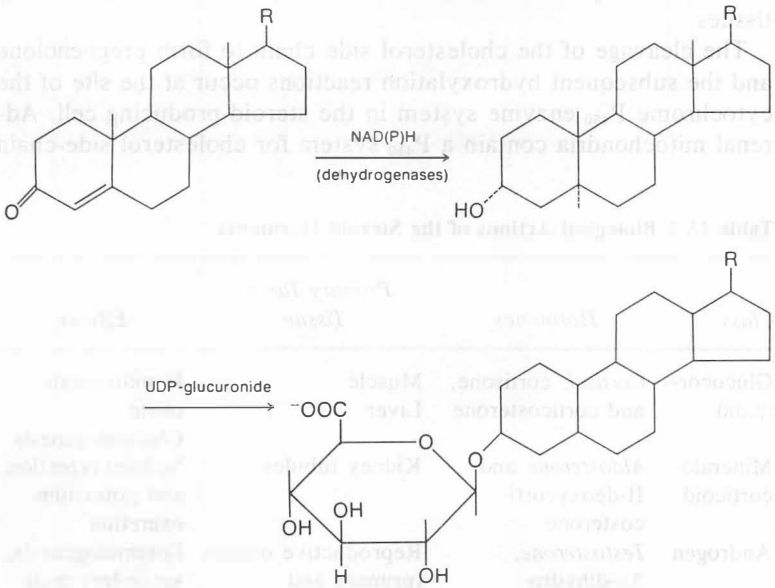
blood. These peripheral tissues include skin, liver, brain, mammary, and adipose tissues. The steroids are not stored in appreciable amounts in the endocrine glands but are secreted into the general circulation and distributed to all tissues of the body. There are two established carrier proteins in blood, cortisol binding globulin (CBG, or transcortin), which binds cortisol, corticosterone, and progesterone; and the sex steroid binding globulin (17β-globulin), which has a high affinity for testosterone and estradiol. Steroids also bind with a low affinity to serum albumin.

Protein binding does not limit the access of the hormone into tissue cells because there is an equilibrium at all times with the free steroid form in solution. At high concentrations of steroids, the concentration of free steroids will increase in blood, as the capacity of the high affinity globulins to bind steroids is exceeded and more steroid is bound to low affinity nonspecific protein sites.

The plasma concentration of the steroid hormones at any one moment in time represents the net difference between the rate of formation and secretion of the hormone by the endocrine gland and the rate of metabolism in liver and excretion by the kidneys. There

appears to be no apparent limit in the capacity of these organs to metabolize steroids, and there is no appreciable storage in the endocrine gland. The rate of turnover of steroid hormones is rapid. The half-life of steroids ranges from 30 to 90 min. The rate of formation and secretion of a steroid hormone by the endocrine gland, therefore, is an essential control point for the biological action of the steroid hormone at its target tissue.

The liver is the primary organ for metabolizing steroid hormones. Reduced steroids are formed by the action of stereospecific dehydrogenases, using pyridine nucleotides as cofactors. The reduced metabolites are conjugated at the hydroxyl groups as sulfates or glucuronides in which they circulate in the blood. They are rapidly excreted into the urine, since kidney clearance is greatest for glucuronide and sulfate conjugates of steroids. Normally only trace amounts of free nonreduced steroids are found in urine; the low clearance is partly a result of their binding to plasma proteins.



The measurement of steroid hormones and their metabolites in blood and in urine is routine at most medical centers. A number of the older classical chemical procedures that have been used in the past have been replaced by newer radioimmunoassay procedures.

Some storage of the estrogenic and progestational steroids occurs in adipose tissue. These steroids also have an appreciable enterohepatic circulation; that is, they rapidly appear in the bile in their conjugated forms and, like cholesterol, enter the gastrointestinal

(GI) tract, and then are reabsorbed via the hepatic portal vein back into the liver system. In contrast, the metabolized androgens and adrenocortical steroids are excreted more rapidly as reduced conjugate forms into the urine. With the reservation concerning some tissue storage for estrogen and progesterone, since total steroid metabolism is relatively rapid, the amount of metabolites of the steroids found in urine can approximate the secretion rate of steroids by the endocrine tissues over the period of time (4–24 h) of urine collection.

Steroidogenesis

The biological actions of the steroid hormones are presented in Table 15.3. There is a common metabolic pathway for the formation of all the steroid hormones initiated by the conversion of cholesterol to pregnenolone. The conversion of pregnenolone to progesterone and to 17α-hydroxy derivatives also occurs in all steroid-producing tissues.

The cleavage of the cholesterol side chain to form pregnenolone and the subsequent hydroxylation reactions occur at the site of the cytochrome P_{450} enzyme system in the steroid-producing cell. Adrenal mitochondria contain a P_{450} system for cholesterol side-chain

Table 15.3 Biological Actions of the Steroid Hormones

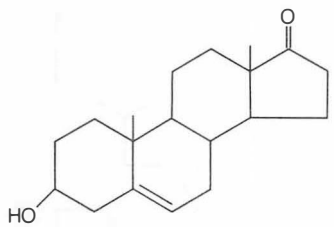
Class	*Hormones*	*Primary Target Tissue*	*Effects*
Glucocorticoid	*Cortisol,* cortisone, and corticosterone	Muscle Liver	Protein catabolism Gluconeogenesis
Mineralocorticoid	*Aldosterone* and II-deoxycorticosterone	Kidney tubules	Sodium retention and potassium excretion
Androgen	*Testosterone,* 5α-dihydrotestosterone, and dehydroepiandrosterone	Reproductive organs (primary and secondary) Muscle	Spermatogenesis, secondary male characteristics, bone maturation, virilization
Estrogen	*Estradiol* and estrone	Reproductive organs (primary and secondary)	Feminization Cyclic rhythms
Progestin	*Progesterone*	Uterus	Nidation and maintenance of pregnancy

cleavage, 11β- and 18-hydroxylation. The endoplasmic reticulum contains a P_{450} system catalyzing 17α-, 21-, and 19-hydroxylation. Steroids are hydroxylated by a microsomal P_{450} system in liver at carbons 2, 7, and 16; however, the major metabolic fate for steroids in the liver is the result of dehydrogenase activity, which leads to the reduced conjugated metabolites of the hormones. Several cytochrome P_{450} systems have been purified with different specificities due to protein differences, which would account for specific hydroxylation sites on the steroid molecule.

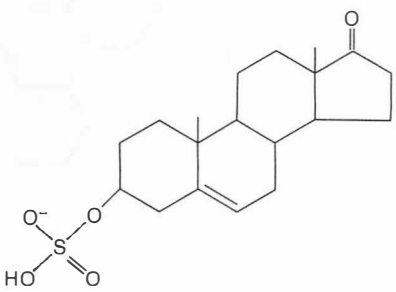
Steroidogenesis is stimulated by pituitary trophic factors, which determine the amount of cholesterol available at the P_{450} enzyme site and the rate of cholesterol side-chain cleavage. The amount of cholesterol converted to the C_{21} steroid pregnenolone is regulated by ACTH in the adrenal cortex and by luteinizing hormone (LH) in the interstitial cells of the testis and ovary. Pregnenolone, once formed, is immediately converted by hydroxylases and dehydrogenases to the final steroid hormone products, which are then released into the circulating plasma.

In the adrenal there are three possible pathways by which the precursors are converted to adrenal steroid hormones (Figure 15.7). Starting with progesterone and 21-hydroxylation, a sequence leads to corticosterone, which is converted in the glomerulosa zone to the final secretory product, *aldosterone*. In a second pathway 17α-hydroxyprogesterone is formed first, followed by 21- and 11β-hydroxylation to give the final secretory product, *cortisol*. Cortisone is not a primary secretory product of the adrenal, but it is formed in metabolism by the oxidation of the 11β-hydroxyl group of cortisol to the 11-ketone. A third pathway in the adrenal involves the 17α-hydroxylation of pregnenolone and subsequent side-chain cleavage by a C_{17-20} desmolase to give the C_{19} compound, *dehydroepiandrosterone* (DHEA). This 17-ketosteroid is a weak androgen, as are its products, 4-androstene-3,17-dione and 11β-hydroxy-4-androstene-3,17-dione. Dehydroepiandrosterone is secreted by the adrenal in large amounts (> 10 mg/day), but its most unusual feature is the fact that it is secreted as a sulfate conjugate. Dehydroepiandrosterone sulfate is present in circulating plasma in relatively large amounts (50–150 μg/dl) and represents a large potential reservoir of precursor steroid for possible conversion to androgen or estrogen in the peripheral tissues.

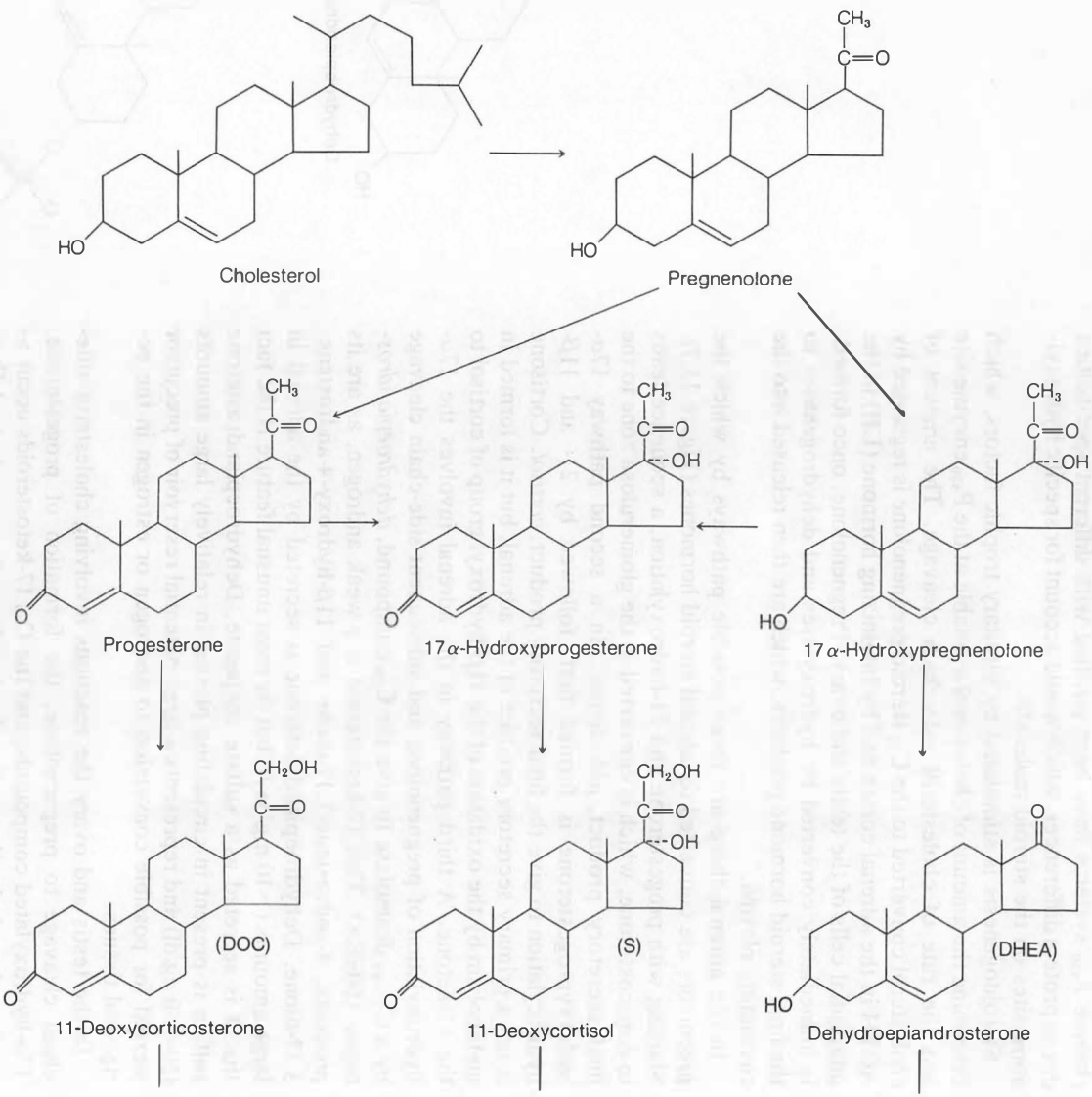
In the testis and ovary the reactions involving cholesterol side-chain cleavage to pregnenolone, the formation of progesterone, 17α-hydroxylated compounds, and the C_{19} 17-ketosteroids occur as in the adrenal. In the testis, the Leydig cell responds to LH to produce the male hormone testosterone (Figure 15.8). In the ovary the production and secretion of progesterone is under the control of



Dehydroepiandrosterone



Dehydroepiandrosterone sulfate


HO
Cholesterol
CH3
C=O
HO
Pregnenolone
CH3
C=O
O
Progesterone
CH3
C=O
---OH
O
17α-Hydroxyprogesterone
CH3
C=O
---OH
HO
17α-Hydroxypregnenolone
CH2OH
C=O
(DOC)
O
11-Deoxycorticosterone
CH2OH
C=O
---OH
(S)
O
11-Deoxycortisol
O
(DHEA)
HO
Dehydroepiandrosterone

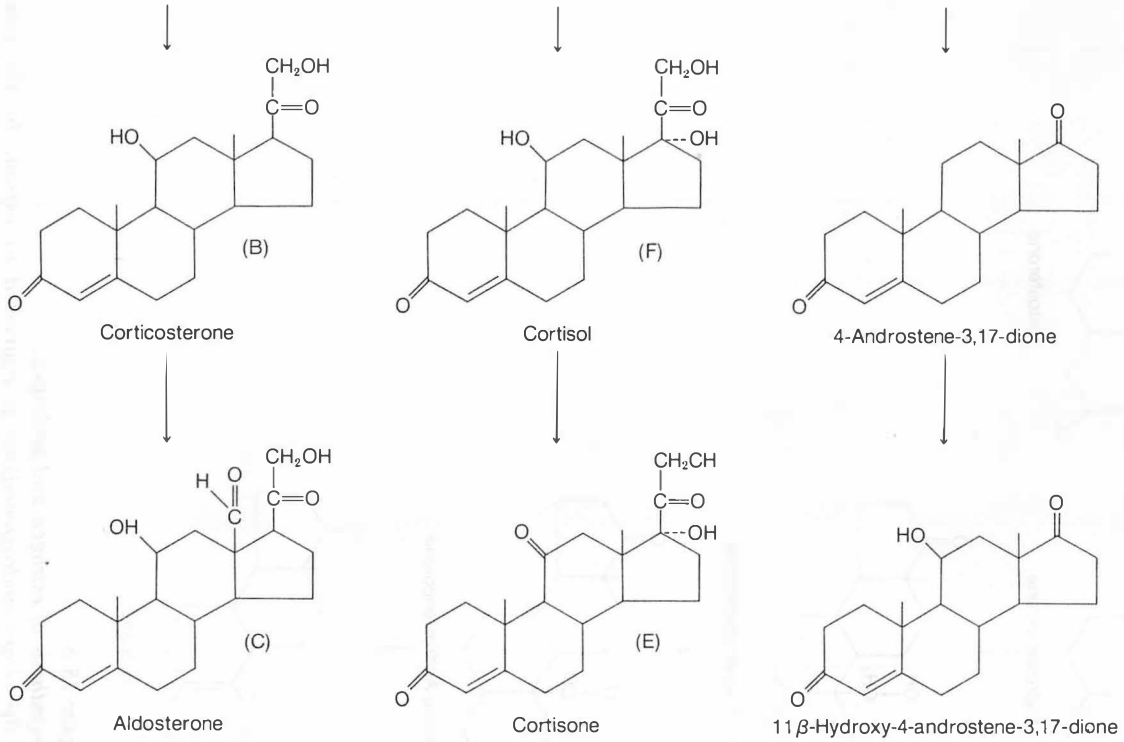


Figure 15.7
Biogenesis of adrenal cortical hormones.

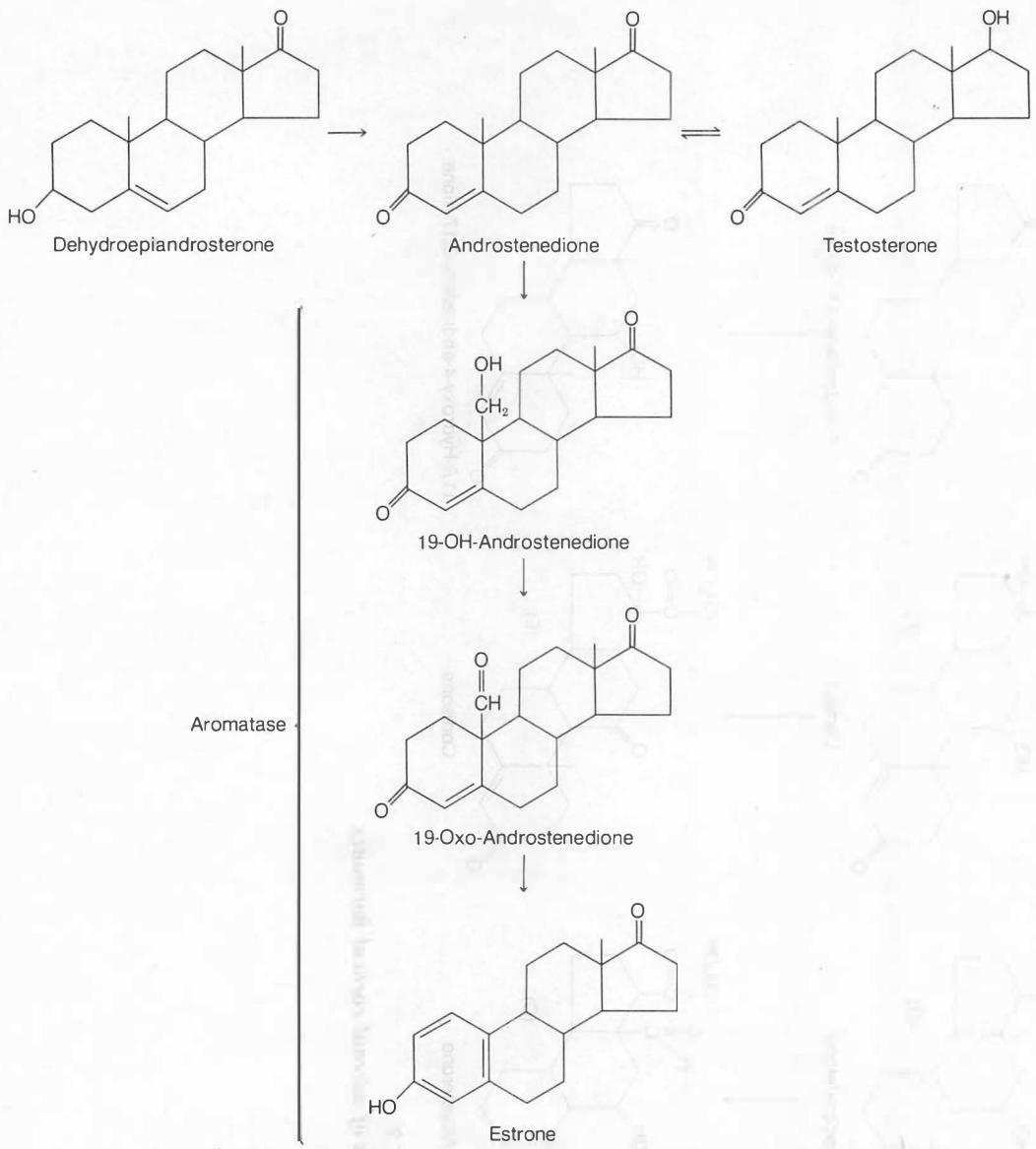


Figure 15.8
Biosynthesis of estrogen and androgen.
In the ovary androstenedione is converted to estrone by the concerted action of an enzyme complex which contains P_{450} hydroxylases for C_{19} and C_2 of the steroid nucleus. The enzyme complex that forms estrogen is referred to as the aromatase enzyme.

LH stimulation of cells of the corpus luteum. Progesterone secretion is cyclic, its production occurs almost entirely during the luteal phase of the menstrual cycle when the corpus luteum is active. During pregnancy, while the corpus luteum persists and after the development of the placenta, larger amounts of progesterone are secreted.

The steroid-producing capacity of the placenta is under trophic hormone control also. The substance is called human chorionic gonadotropin (HCG), a protein with LH activity but with an amino acid sequence different from that of LH and which is immunologically distinct from LH. As the name implies, HCG is not derived from the anterior pituitary but is produced by cells of the placenta. During pregnancy, steroid sulfates formed in fetal tissues are converted by the placenta to estrogens. A sulfatase converts dehydroepiandrosterone sulfate to the free steroid, which is then converted to estrone by the enzyme complex aromatase (Figure 15.8). By a similar sequence, 16α-hydroxydehydroepiandrosterone sulfate formed in the fetus is transported to the placenta, where it is hydrolyzed by the sulfatase and aromatized to the estrogen estriol. Estriol is the major estrogen metabolite formed in late pregnancy.

Regulation of Steroidogenesis

The dynamic state that controls secretion of steroid hormones is the result of the interaction of a number of regulatory factors. For the adrenal this includes the steroid cortisol, which exerts a negative feedback effect on the hypothalamic–pituitary system; the pituitary peptide adrenocorticotropic hormone (ACTH), which stimulates the adrenal cell; a hypothalamic factor, corticotropin releasing factor (CRF), which stimulates the pituitary cell to secrete ACTH; and other neurotransmitter agents that regulate the secretion of CRF.

The hypothalamus serves as an integrating center receiving signals from the central nervous system and the higher cortical centers via neurotransmitters. The composite signal results in the secretion of CRF into the portal blood system draining the hypothalamus, which then enters the pituitary to stimulate ACTH secretion. Although CRF was the first of the hypothalamic releasing factors to be demonstrated, its chemical structure is not known, and it is not available for clinical use.

In response to CRF, the pituitary cells secrete ACTH. ACTH is available for clinical use as the natural polypeptide containing 39 amino acids and as a totally synthetic polypeptide containing the first 24 amino acids with complete biological activity. ACTH binds to the plasma membrane of the cells of the fasciculata-reticularis zone in the adrenal cortex to initiate a chain of events leading to the secretion of steroid hormones (Figure 15.9).

As a consequence of ACTH binding there is an activation of adenylate cyclase to convert ATP to cAMP. With increasing doses of ACTH, there is a corresponding increase in the cellular concentration of cAMP. cAMP activates protein kinase, which phosphorylates phosphorylase, and a cholesterol esterase, which hydrolyzes cholesterol esters stored in lipid droplets to the free cholesterol form. In particular, there is enhanced uptake of cholesterol from the lipoprotein–cholesterol complex circulating in plasma. The formation of cholesterol de novo by the adrenal HMG-CoA reductase system (Chapter 10) is increased as well. The free cytoplasmic cholesterol probably is transported by a carrier protein and converted by the mitochondrial cytochrome P_{450} enzyme system to pregnenolone.

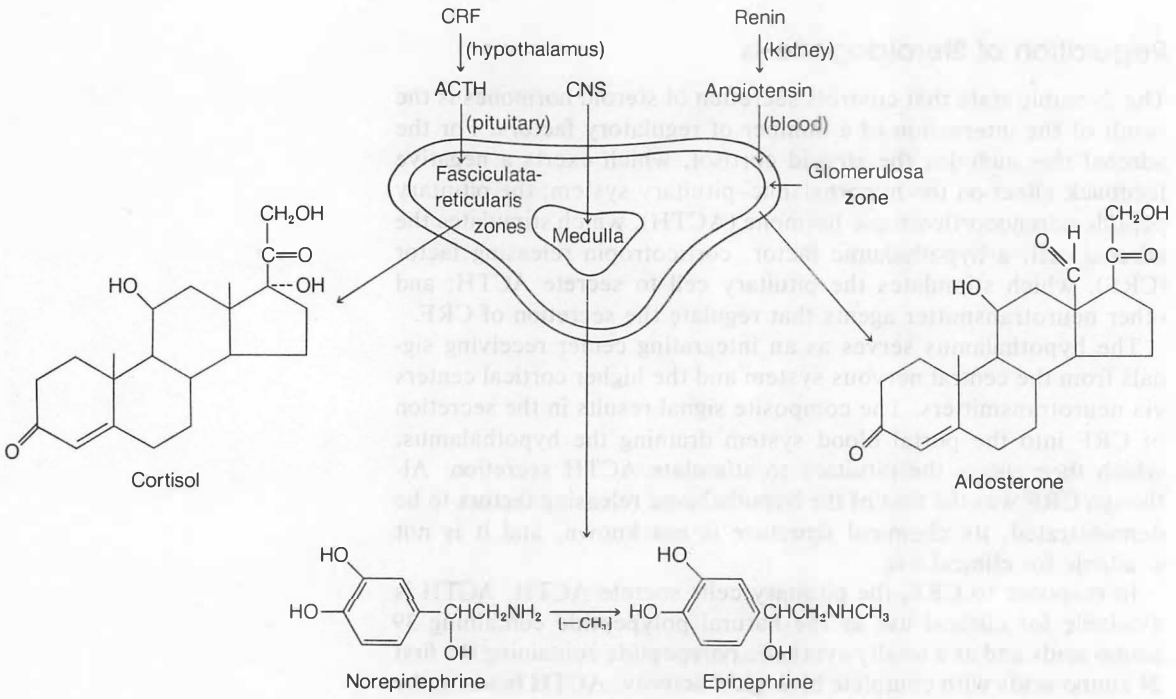


Figure 15.9
Functional zonation of the adrenal gland.

Functional Zonation of the Adrenal Gland

The adrenal gland can be differentiated by structure and function into three independent biologically active zones: (1) the centrally located *medulla,* (2) the cortex layers of the *fasciculata-reticularis* zones, and (3) a thin outer cortex layer of the *glomerulosa* zone beneath the adrenal capsule (Figure 15.9).

The adrenal medulla, which consists essentially of nerve tissue (chromaffin), responds to splanchnic nerve stimulation or hypoglycemia by increased formation of norepinephrine and epinephrine. The enzyme *N*-methyltransferase, with *S*-adenosylmethionine as the methyl donor, converts norepinephrine to epinephrine. Epinephrine and norepinephrine are secreted by the adrenal medulla in a ratio of ~4 : 1 in the human. Despite the close proximity of the medullary cells to the cortical cells producing steroid hormones, there appears to be little physiological interaction of the catecholamines on the cortex, or for action of the steroid hormones on the medulla. There is a prominant circadian biorhythm in the enzyme activities of the medulla and cortex, which can be entrained by light/dark and activity schedules. The *N*-methyltransferase activity to form epinephrine in the medulla has been related to the cortisol circadian cycle.

The cells of the fasciculata-reticularis zone of the adrenal cortex appear to comprise a continuum morphologically, the cells of the fasciculata being somewhat larger, containing more lipid droplets, and arranged in orderly columns; however, there is very little evidence that the fasciculata cells and the reticularis cells are functionally different. It has been suggested that the fasciculata-reticularis zone contains cells that produce the C_{19} 17-ketosteroids of the adrenal. It is clear, however, that both zones respond to ACTH stimulation; the loss in lipid droplets in the fasciculata zone with stimulation is particularly prominent. The hypertrophy of the adrenal cortex as a result of chronic ACTH stimulation is confined to the fasciculata-reticularis zones. The prominent biochemical feature of ACTH stimulation is enhanced steroid production featuring cortisol secretion. There is increased secretion of dehydroepiandrosterone and lesser amounts of other C_{21} and C_{19} steroids as well.

The glomerulosa (capsular) zone consists of a thin layer of cells beneath the outer capsule of the adrenal gland. The most prominent biochemical feature of this zone is the production of the mineralcorticoid aldosterone. The C_{18}-aldehyde group unique for aldosterone is derived from C_{18}-hydroxylated steroids in the glomerulosa zone. Aldosterone secretion is stimulated by sodium restriction or potassium excess, or a combination of both. The width and number of cells of the glomerulosa zone is increased with chronic sodium restriction.

Angiotensin I (10 amino acids)

Asp^{1}-Arg^{2}-Val^{3}-Tyr^{4}-Ile^{5}-His^{6}-Pro^{7}-Phe^{8}-His^{9}-Leu^{10}

↓ (converting factor)

Angiotensin II (8 amino acids)

Asp-Arg-Val-Tyr-Ile-His-Pro-Phe

↓

[des-Asp^{1}]-Angiotensin II (7 amino acids)

Arg-Val-Tyr-Ile-His-Pro-Phe

Angiotensin II, a peptide containing eight amino acids, plays an important role in the regulation of aldosterone formation and secretion. The regulatory chain of events is initiated in the kidney with the secretion of renin. Renin, a kidney enzyme secreted in response to osmotic changes, anoxia, or kidney trauma, acts upon a blood precursor protein, angiotensinogen. A 10-amino acid peptide, angiotensin I, which has potent blood pressor activity, is formed. The decapeptide gives rise to the octapeptide angiotensin II by an enzyme converting factor. An active heptapeptide is also formed. Participating agents in the regulatory system comprise aldosterone, angiotensin II, and sodium, among others, but their interactions at the molecular level have been difficult to demonstrate.

Biological Actions of Adrenal Steroids

The biological actions of corticosteroids are classified in terms of three major categories: mineralocorticoid activity, glucocorticoid activity, and antiinflammatory activity (Table 15.4).

Table 15.4 Biological Actions of Corticosteroids (Cortisone = 1.0)

Steroid	*Na^{+} Retention*	*CHO Activity*[a]	*Antiinflammatory*
Cortisone	1.0	1.0	1.0
Cortisol	1.5	1.5	12.5
Corticosterone	2.5	0.5	0
11-Deoxycorticosterone	30.0	0	0
Aldosterone	600.0	0.3	0

[a] Effect on carbohydrate metabolism.

Mineralocorticoid Activity

The most active mineralocorticoid and the physiologically effective agent produced by the adrenal is aldosterone. The action of aldosterone is at the distal convoluted tubule of the kidney to promote Na^+ reabsorption. The Na^+ retention is accompanied by a corresponding excretion of K^+ and H^+. In the absence of aldosterone (as in Addison's disease) there is no mechanism to retain Na^+, and the body will continue to lose Na^+, even with severe Na^+ restriction in the diet. The excretion of Na^+ (and the consequent retention of K^+ and H^+ ions) is a critical factor in adrenal insufficiency or after adrenalectomy.

Glucocorticoid Activity

Cortisol is the major glucocorticoid secreted by the human adrenal cortex. The term glucocorticoid refers to the activity of steroids that can affect carbohydrate metabolism. Specifically, the activity relates primarily to the ability of steroids to increase blood glucose levels, and second, to the increased deposition of glycogen in the liver. Both effects are due to enhanced liver gluconeogenesis as a consequence of steroid actions discussed below.

Glucose utilization by the muscle cell is inhibited by steroid activity (Figure 15.10). This action opposes the insulin effect to increase muscle permeability to glucose and amino acids. The antianabolic action of the steroid results in a breakdown (catabolic action) of muscle protein, a consequence of protein turnover.

Cortisol receptor sites have been demonstrated in the nuclei of liver cells. Increased induction of the enzymes involved in transamination and in gluconeogenesis have been found in liver after cortisol administration (Figure 15.11). As a result of cortisol activity, the

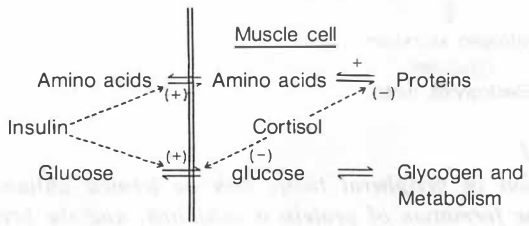


Figure 15.10
Insulin as an anabolic agent enhances (+) the incorporation of glucose and amino acids into muscle cells. Cortisol decreases (−) glucose uptake and incorporation of amino acids into protein, which leads to increased protein breakdown (catabolism).

CLIN. CORR. 15.1 BIOLOGICAL ACTIONS OF GLUCOCORTICOIDS

The term *permissive action* is used frequently to describe the physiologic role of the glucocorticoids. That is, they act in a synergistic manner with other agents to allow certain events to occur. It is not possible to define these actions at the molecular level, however, since those effects that can readily be observed in vivo cannot be reproduced in an in vitro system. This is an indication that the observed effects are the result of an indirect rather than a direct action of the steroids.

The glucocorticoids, cortisol, and the synthetic analogs of cortisol are used extensively in medicine as *antiinflammatory* agents. Since they can diminish normal but undesirable body responses to noxious stimuli or trauma, advantages are gained by their use in counteracting stressful situations and in decreasing pain and discomfort. In this respect, they are used as *palliative* agents, as the symptoms rather than the cause of the disability or disease is treated. The danger of *iatrogenous* sequellae, a result of the side effects of steroid treatment, often will set a limit on the advantages achieved by their use.

amino acids released from peripheral tissues are converted by transaminase to keto acids and via the gluconeogenic pathway to increased formation of glucose (Chapter 7). Part of this glucose is released by the liver to increase blood glucose, and part is utilized by the liver cell to form glycogen. The glycogen deposition occurs by the action of insulin on glycogen synthetase. The enzymes involved in these reactions, transaminases, fructose 1,6-bisphosphatase, and glucose 6-phosphatase, have been shown to be increased as a result of cortisol activity. Note that as a result of cortisol action, there is an increase in protein synthesis in the liver while protein breakdown to amino acids is occurring in peripheral muscle tissues (Clin. Corr. 15.1).

Antiinflammatory Activity

Antiinflammatory activity refers to those actions of cortisol that interfere with the normal cellular processes of inflammation, wound

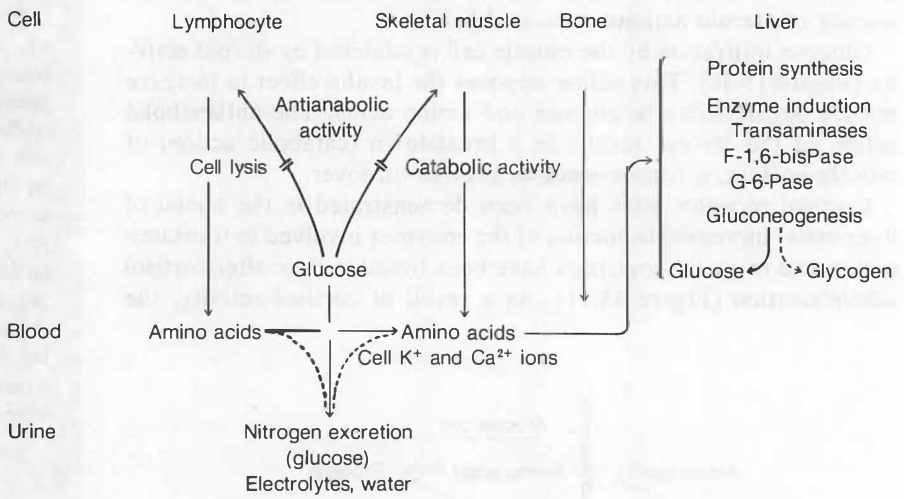


Figure 15.11

Cortisol action in peripheral tissue can be termed antianabolic and catabolic, the formation of protein is inhibited, and the breakdown of protein to amino acids is enhanced.

This activity is seen in lymphocytes, skeletal muscle, and bone. In liver, cortisol induces the synthesis of transaminase and enzymes involved in gluconeogenesis (fructose 1,6-bisphosphatase and glucose 6-phosphatase), leading to the formation of glucose and liver glycogen.

healing, and growth. There is a marked specificity for this activity shared only by cortisol and cortisol derivatives (Table 15.4). There are now a large number of synthetic cortisol analogs available for clinical use. These orally active steroids have greater potency than cortisol with respect to antiinflammatory activity and glucocorticoid activity, but have diminished mineralocorticoid activity.

It is on the basis of their antiinflammatory activity that the glucocorticoids have attained such widespread use as palliative agents in clinical therapy. These uses include treatment for asthma and allergies; inhibition of immune processes and tissue rejection, as in organ transplants; increasing resistence to stress and life-threatening situations; and a wide variety of disease processes.

15.4 STEROID SEX HORMONES: REGULATION AND BIOLOGICAL ACTIVITY

The major function of the steroid sex hormones is the development, growth, maintenance, and regulation of the reproductive system. Sex steroids are classified according to biological activity: *androgens* are male sex hormones and are C_{19} steroids; *estrogens* are female sex hormones and are C_{18} steroids (ring A is phenolic and lacks the C_{19}-methyl); and *progesterone* is a C_{21} steroid which is secreted during the luteal phase of the ovarian cycle and during pregnancy.

Androgens

Androgens are produced by the Leydig cells of the testes and the adrenals in both sexes. The ovary also produces androgens in small amounts. Testosterone is the major androgen produced by the testes. About 6–7 mg/day are secreted in the adult male. Testosterone is synthesized from cholesterol by a pathway shown in the section on steroidogenesis (Figure 15.8). Although the main secretory product of the testes is testosterone, the active hormone in many tissues is not testosterone, but rather a metabolite, 5α-dihydrotestosterone (DHT).

The Leydig cells are the principal site of production of androgens made in the testes. The Sertoli cells are capable of forming androgen and also some estrogen. One function of Sertoli cells appears to be the production of androgen binding protein (ABP), which is secreted into the seminiferous tubules. This protein specifically binds testosterone and DHT. Its presence in the seminiferous tubules is asso-

ciated with androgen action on spermatogenesis. The production of androgen is primarily under regulation of LH, while FSH (follicle stimulating hormone) acts on the Sertoli cell to stimulate production of ABP and other factors, including, possibly, estradiol.

Androgens influence the development, maintenance, and function of the male reproductive organs and the male secondary sex characteristics. Androgens have widespread effects throughout the body as well as localized effects on specific tissues. These include the structures concerned with the formation and delivery of sperm, the low pitch of the male voice due to enlargement and thickening of the vocal cords, the male pattern of hair distribution. They have profound anabolic activity, leading to the retention of nitrogen and increased muscle and bone mass. They are responsible for bone maturation and the cessation of growth after puberty. There is an effect on the brain, leading to characteristic male sexual behavior and aggressiveness. With respect to these biological actions, receptor sites for androgens have been demonstrated in brain, muscle, and other target tissues where androgen effects are known to occur.

Regulation

Receptor binding of LH by the Leydig cell leads to the stimulation of membrane-bound adenylate cyclase and the production of cAMP. As is the case for ACTH and the adrenal cell, the result of cell stimulation by LH is increased cleavage of the side chain of cholesterol to form pregnenolone. Pregnenolone is immediately converted to testosterone by P_{450} enzymes of the Leydig cell endoplasmic reticulum (Figure 15.8). An increased secretion of testosterone is seen within a few minutes after the administration of the gonadotropins LH and FSH.

Metabolism of Androgen

Both testosterone and DHT circulate in blood bound to 17β-globulin, or testosterone-estrogen binding globulin (TeBG). The liver is the major site of metabolism of testosterone. The main products of metabolism are conjugated and excreted via the kidney as sulfates and glucuronides. Figure 15.12 illustrates the conjugate structures for androsterone (*a*), a weak androgen, and etiocholanolone (*b*), an isomer of androsterone devoid of androgenic activity. Together with dehydroisoandrosterone sulfate (*c*), they comprise the principal steroids referred to as urinary 17-ketosteroids (17-KS). However, the androgens derived from the adrenal (Figure 15.7) are also precursors of the urinary 17-KS, and are secreted in amounts greater than is testosterone and other androgens derived from the ovary or testis. For this reason, urinary 17-KS reflect adrenal function to a greater extent than gonadal function. Urinary values of 17-KS are

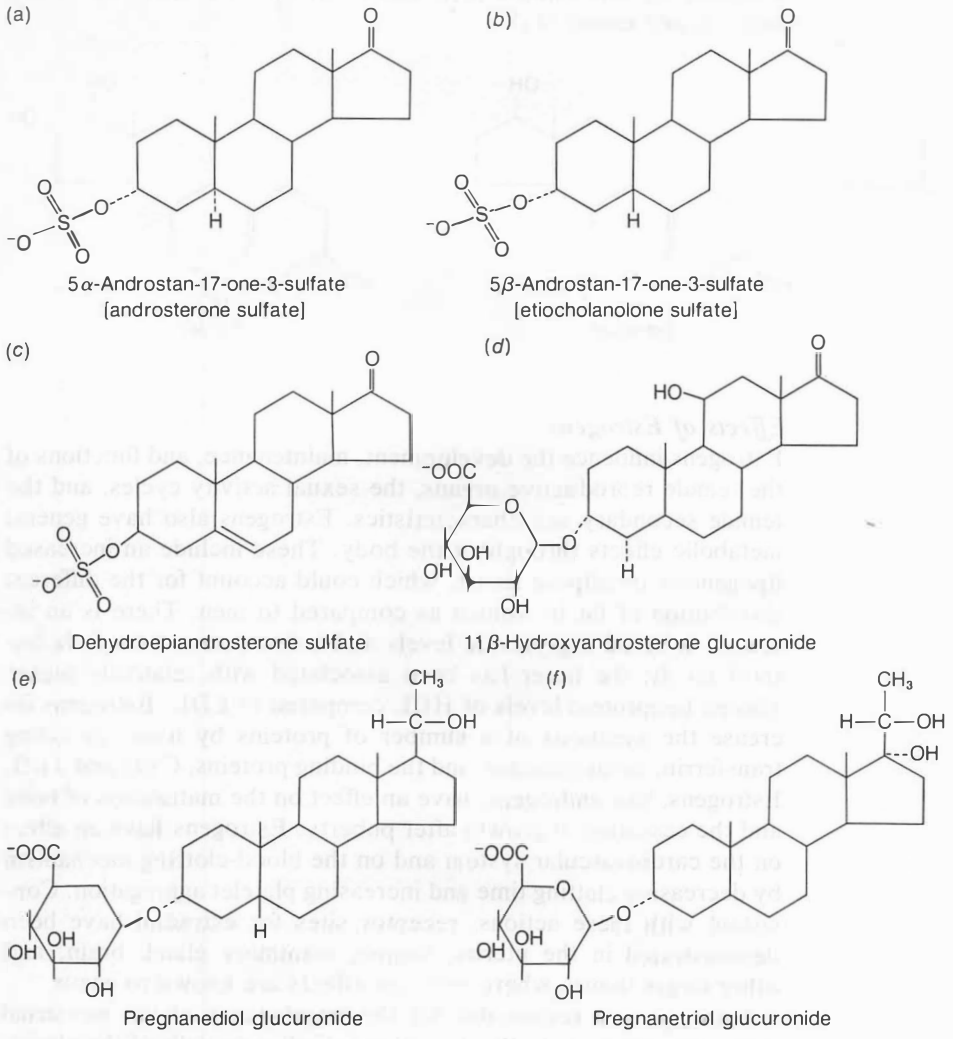
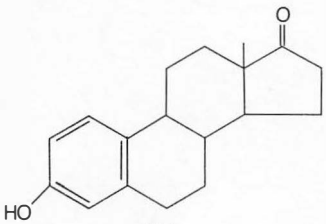


Figure 15.12
Steroid metabolites in urine.

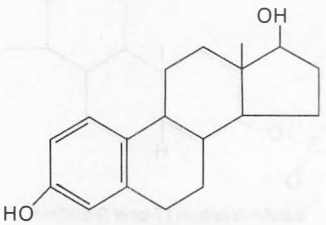
approximately one-third higher in males than females due to the testicular production of testosterone.

Estrogens

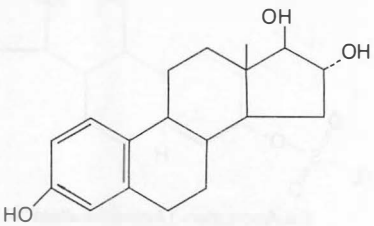
Estrogens are produced by the follicles and corpus luteum of the ovary and by the placenta during the second and third trimesters of pregnancy in greatly increased amounts. The ovary secretes both estradiol (E_2) and estrone (E_1), whereas the placenta produces these steroids and estriol (E_3).



(E_1)
Estrone



(E_2)
Estradiol



(E_3)
Estriol

Effects of Estrogens

Estrogens influence the development, maintenance, and functions of the female reproductive organs, the sexual activity cycles, and the female secondary sex characteristics. Estrogens also have general metabolic effects throughout the body. These include an increased lipogenesis in adipose tissue, which could account for the different distribution of fat in women as compared to men. There is an increase in blood triglyceride levels and a decrease in blood cholesterol levels; the latter has been associated with relatively higher plasma lipoprotein levels of HDL compared to LDL. Estrogens increase the synthesis of a number of proteins by liver, including transferrin, ceruloplasmin, and the binding proteins, CBG and TBG. Estrogens, like androgens, have an effect on the maturation of bone and the cessation of growth after puberty. Estrogens have an effect on the cardiovascular system and on the blood-clotting mechanism by decreasing clotting time and increasing platelet aggregation. Consistent with these actions, receptor sites for estradiol have been demonstrated in the uterus, vagina, mammary gland, brain, and other target tissues where estrogen effects are known to occur.

Estrogens are responsible for the maintenance of the menstrual cycle. They have growth-promoting activities on cells of the uterus, vagina, Graafian follicles of the ovary, and the mammary gland. Estrogens induces in the uterus and mammary gland the synthesis of

progesterone receptors which are necessary for these tissues to respond to progesterone. This activity of estrogen prior to progesterone action is referred to as "estrogen priming" and is an essential feature in the growth and function of these target tissues.

Regulation

Receptors for LH have been demonstrated in the plasma membranes of the interstitial cells in the stroma of the ovary and in the corpus luteum of the ovary. The interstitial cells are stimulated by LH and the granulosa cells of the Graafian follicle are stimulated by FSH. The combined action of these two systems are required for the synthesis of estradiol.

Metabolism of Estrogens

Estradiol is bound to 17β-globulin in the circulating plasma. Estrone and estriol are only weakly bound. Estradiol undergoes a complex metabolism in peripheral tissues and in the liver. Metabolic events include a reversible oxidation to estrone and irreversible hydroxylation at C-2 and C-16. The metabolites are conjugated with sulfuric or glucuronic acid and excreted either into urine or bile. Biliary metabolites may undergo further metabolism by action of the intestinal flora and are reabsorbed into the portal circulation.

A metabolite of estradiol, which comprises at least 20% of the total amount secreted in human, is the 2-hydroxyl derivative (2,3-diol in ring A) (Figure 15.13). These derivatives, referred to as

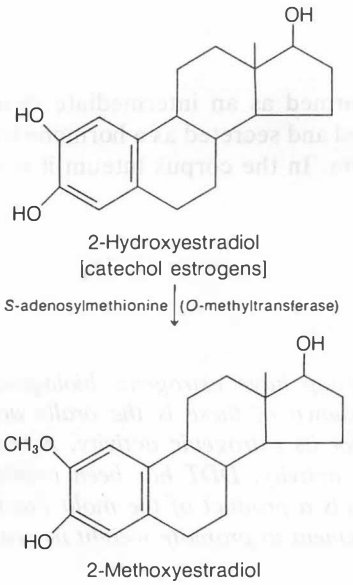
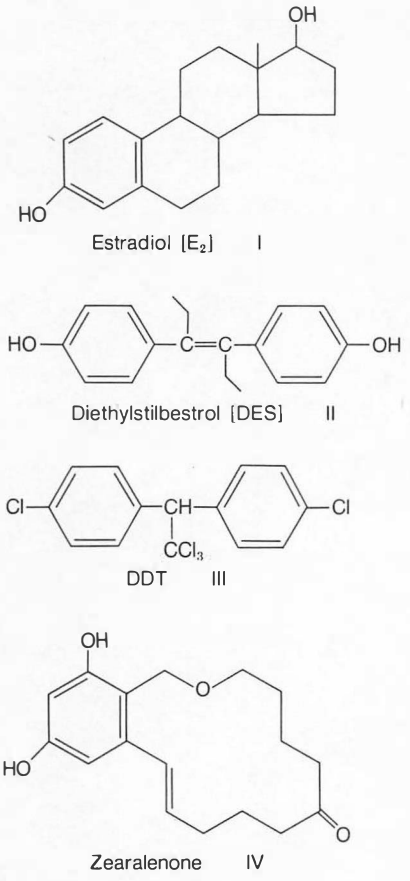


Figure 15.13
Formation of 2-methoxyestradiol.



catechol estrogens, are unstable in alkali. Since alkali is ordinarily used in extraction for estrogen assays, the assays for catechol estrogens require special procedures.

The conversion to catechol estrogen occurs in a number of tissues, including the brain. A portion of estrogen is irreversibly (covalently) bound to tissue protein, and it has been proposed that these estrogen forms are involved in their expression of biological activity. It is also of interest that the same enzyme, *O*-methyltransferase, which converts the 2-hydroxyl to the 2-methoxyl derivative (Figure 15.13), also forms the methoxyl derivative of epinephrine and norepinephrine. An interaction and control of epinephrine action by estrogen, which is known to occur, thus may be exerted through the process of altering the rate of inactivation of these substrates by competition for the *O*-methyltransferase.

Estrogens have assumed an important role in environmental pollution. There are a number of aromatic substances found in plants (phytoestrogens) or produced in industry (insecticides), which may have important consequences on animal and human ecology (Figure 15.14). By affecting bird eggshell quality (as with DDT) or reproductive activity of grazing animals due to their intrinsic estrogenic activity, many of these substances of synthetic origin have added to the complexity of the important problem of waste disposal. In this context, with respect to possible adverse affects for life on this planet, the student might deliberate on the fortunate circumstance that natural steroids such as estradiol serve as substrates for microbial degradation in the soil.

Progesterone

Progesterone is formed as an intermediate in all steroid-secreting cells. It is produced and secreted as a hormone by the corpus luteum and by the placenta. In the corpus luteum it is derived from blood

Figure 15.14
Estrogenic substances.
A wide variety of chemical compounds containing a phenolic group have estrogenic biological activity, which is comparable to the natural estrogen, estradiol (E_2, I). The best known of these is the orally active synthetic compound diethylstilbestrol (DES, II) which has been used clinically for its estrogenic activity. A number of insecticides that have a phenolic group, such as DDT (III), have estrogenic activity. DDT has been implicated in the quality and fragility of eggs of domestic and wild birds. Zearalenone (IV) is a product of the mold Fusarium, which occurs in grain. It has been used as a substitute for DES, as a food supplement to promote weight increase in livestock.

and ovarian cholesterol (Figure 15.7). During pregnancy maternal circulating cholesterol is the chief precursor to progesterone in the placenta, and near term the placenta secretes 200–300 mg progesterone/day. Production of progesterone by the corpus luteum is about 20–40 mg/day and is under the control of LH.

Biological Actions of Progesterone

Progesterone is necessary for the implantation of the fertilized ovum and for the inhibition of uterine contraction in the maintenance of pregnancy. Progesterone stimulates the growth of the secretory glandular portions of the uterine and breast tissue after estrogen priming. It has a thermogenic effect of unknown origin which raises the normal body temperature by 0.4–1.4°F. Both the thermogenic effect and the large increment in progesterone secretion by the corpus luteum define the luteal phase of the menstrual cycle and are used most commonly as indicators that ovulation has occurred.

Progesterone as well as estradiol, in relatively large amounts, will inhibit ovulation by suppression of gonadotropic stimulation. Orally active, long-acting synthetic derivatives of progesterone and estradiol also will prevent ovulation in relatively small amounts. These are the active agents in the contraceptive pill.

Metabolism of Progesterone

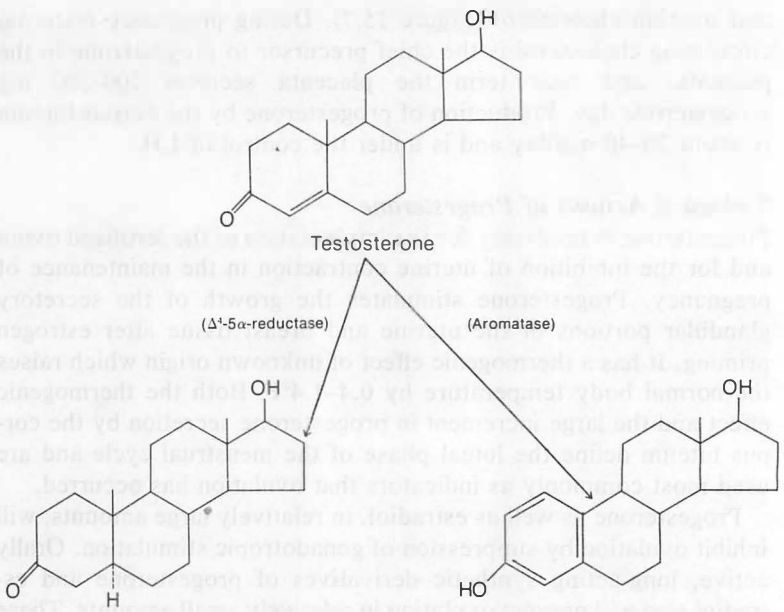
Progesterone circulates in plasma bound to transcortin. Transcortin (CBG) actually has a higher affinity for progesterone than for cortisol. The major urinary metabolite of progesterone is pregnanediol glucuronide (Figure 15.12). Progesterone and its metabolites also have an extensive enterohepatic circulation. Urinary excretion of pregnanediol glucuronide is used as an index of production of progesterone by the corpus luteum and during pregnancy. Pregnanetriol glucuronide is a urinary metabolite of 17α-hydroxyprogesterone, which is increased in certain types of adrenal abnormality.

Peripheral Metabolism of Androgens and Estrogens

Many tissues in the body, in addition to the endocrine glands, liver, and kidney, contain enzymes that metabolize steroids. These enzymes are primarily dehydrogenases and hydroxylases. The conversion of testosterone to 5α-dihydrotestosterone (5α-DHT), for example, is by the action of a Δ^4-5α-reductase. The conversion of testosterone or 4-androstene-3,17-dione to estrogen is due to the action of the enzyme complex aromatase. These reactions occur in

CLIN. CORR. 15.2 ESTROGEN RECEPTORS IN BREAST CANCER

Hormone treatment and endocrine ablation for the treatment of prostate and breast cancer was first introduced in the 1940s by Charles Huggins at the University of Chicago. A further advance in the clinical evaluation for the treatment of breast cancer was the application of a binding assay for estrogen-receptor interaction developed by Elwood Jensen and co-workers at the University of Chicago in the 1960s. It was reasoned that hormone-dependent tumors, those that respond to estrogen for growth, would have demonstrable cytoplasmic and nuclear receptors for estrogen. A positive binding assay using tritium-labeled estradiol and extracts of tumor biopsy tissue should differentiate those tumors that would be most likely to respond to subsequent endocrine ablation and hormone therapy. In very few instances is there a response as a result of such treatment when the estrogen-receptor assay is negative, as compared to a response in 50% of the cases that demonstrate the presence of estrogen receptors in the tumor cells.



tissues such as skin, adipose tissue, and brain. Note that the two reactions forming 5α-DHT, an androgen, and estradiol, an estrogen, are irreversible. Testosterone, the principal steroid sex hormone secreted by the testes, may be considered a prohormone, since 5α-DHT is a biologically active androgen and binds avidly to testosterone receptors. 5α-DHT and estradiol formation should be considered as two distinct and essential features of testosterone biological activity. The role of steroids on the CNS and brain tissue in sexual maturation and behavior and the role of steroids in the etiology of mammary carcinoma are two examples of clinical interest in which metabolic alterations by peripheral tissues have important consequences (Figure 15.15). (See Clin. Corr. 15.2.)

15.5 THYROID HORMONES

Formation of T_4 and T_3

There are two major active hormones produced and secreted by the follicle cells of the thyroid gland, thyroxine (T_4) and triiodothyronine (T_3) (Figure 15.16). With respect to biological activity they are simi-

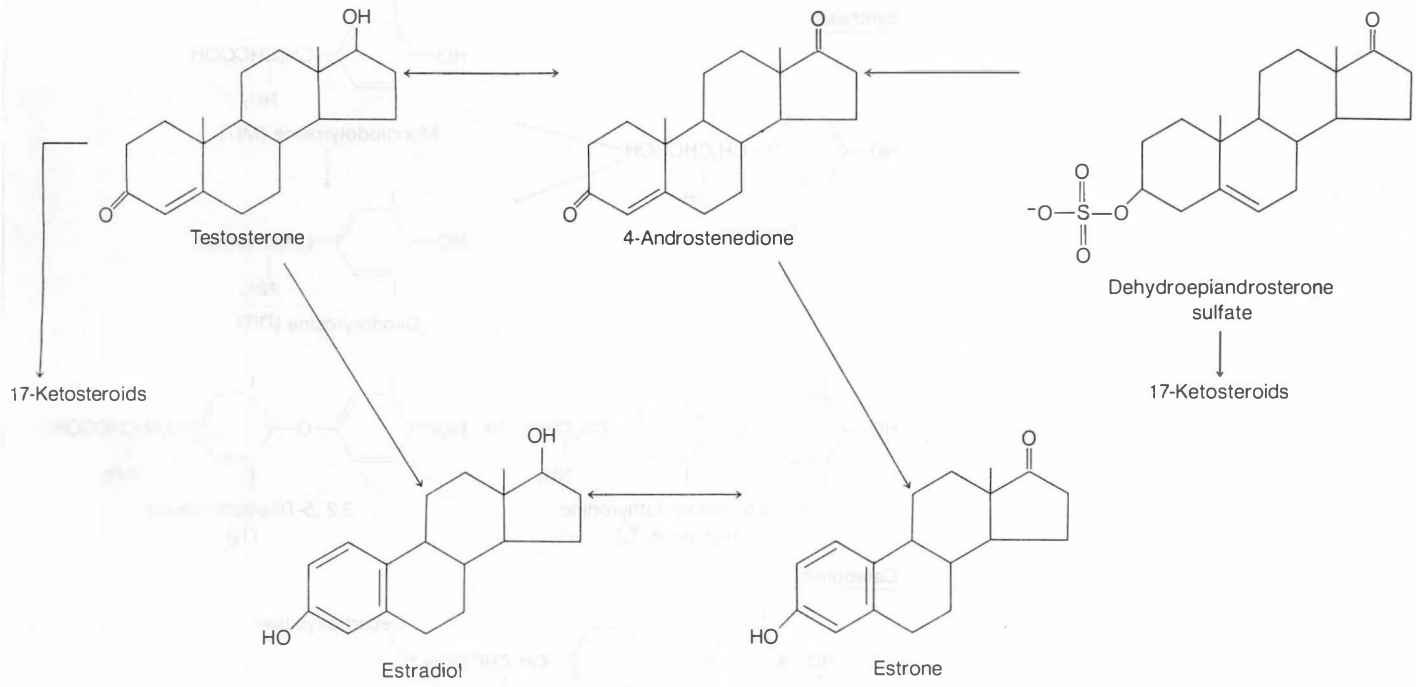


Figure 15.15
Interaction of biologically active steroids in peripheral tissues.

lar qualitatively, but there are important differences. Of the two, T_3 is considerably more active, and it is thought that T_4 activity is due to its conversion to T_3. On a weight basis, T_3 is 3 to 5 times more active than T_4, depending upon the bioassay used. The half-life of the thyroid hormones in the body is considerably longer than for any of the other hormones for reasons not well understood: $T_{1/2}$ for T_4 is 6 days and $T_{1/2}$ for T_3 is 2 days. Binding to plasma thyroxine binding globulin (TBG) and other proteins is extensive with ~0.04% of T_4 and T_3 in the free form. Metabolic clearance could account for the differences in half-life partially, but other factors must be involved. As with cortisol-binding globulin and circulating cortisol (about 95% bound), the free hormones, which are in equilibrium with the bound forms, have unhindered access to the tissue cells. Other than as a storage function the physiological role for TBG (and for CBG) remains to be established. The high plasma protein-bound levels of active hormone could be related to a decrease in kidney clearance and liver inactivation.

More T_4 is produced by the thyroid than T_3, and plasma levels for T_4 (4.5–13.0 μg/dl) are considerably higher than for T_3 (0.06–0.20

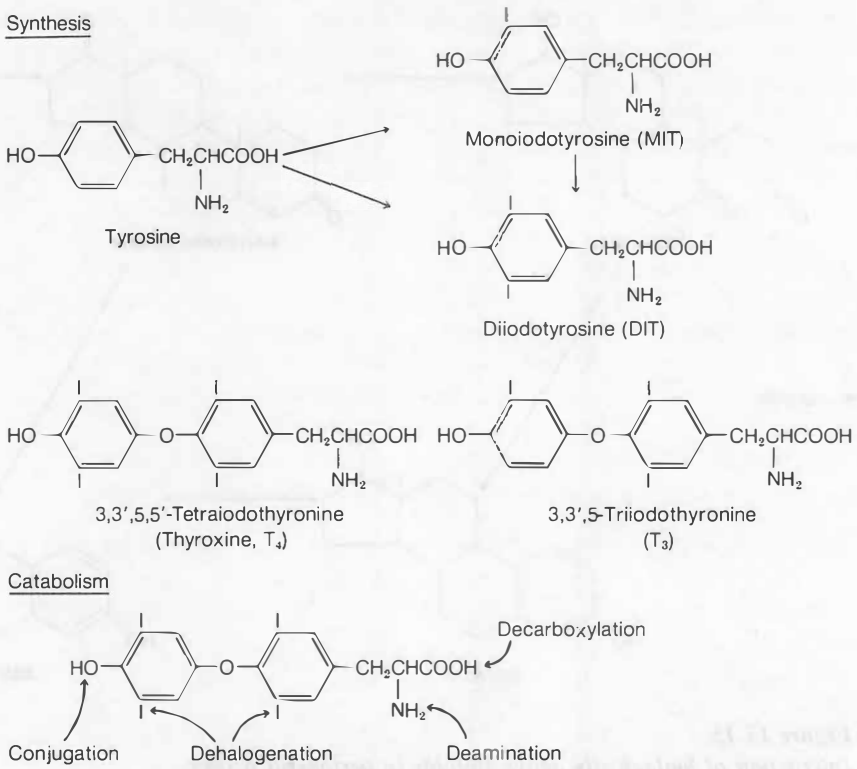


Figure 15.16
Thyroid hormone metabolism.

μg/dl). A unique feature of the thyroid hormones is the presence of organically bound iodine. Other proteins may contain small amounts of iodine, and the tyrosine precursors contain some iodine; however, the bulk of the halide is maintained very efficiently in the thyroid tissue in the form of T_4 and T_3. Normal iodine plasma levels are less than 1 μg/dl and the small loss of iodide (75–100 μg/day) is usually adequately replaced in the diet, which ordinarily contains more than 200 μg/day.

Regulation of Thyroid Hormone Formation

Regulation of thyroid hormone formation by the thyroid follicle is under the control of the pituitary hormone, thyroid stimulating hormone (thyrotropin, TSH). TSH formation and secretion is regulated by the hypothalamic *t*hyrotropin *r*eleasing *f*actor, TRF. TRF contains only three amino acids, pyroglutamyl-histidyl-prolinamide and

was the first of the releasing factors identified and synthesized. Feedback inhibition by both the more active T_3 and by T_4 is exerted at the hypothalamic and pituitary level; however, greater inhibitory effect occurs at the pituitary level. Since the original isolation, TRF has been found in other tissues, and the release of prolactin as well as TSH occurs with TRF administration. This raises questions about other biological roles for TRF, as a neurotransmitter, for example, but its function as a releasing agent for TSH as a true physiological event is generally accepted.

The accumulation of I^- ion in the thyroid follicle exceeds that of any other tissue. The extensive uptake of I^- is referred to as the "iodide trap" and is a function of TSH activity (Figure 15.17). Iodide is oxidized by a thyroid peroxidase, resulting in organic binding to tyrosine molecules. First, a monoiodo- and then a diiodotyrosine is formed. Then the coupling of two iodotyrosines results in the formation of tetraiodo- and triiodothyronine molecules. These iodinated compounds are linked covalently in a peptide linkage to a large (mol wt = 700,000) protein molecule called thyroglobulin. The reactions leading to the formation of the iodinated thyroglobulin in the colloid of the thyroid follicle are controlled by

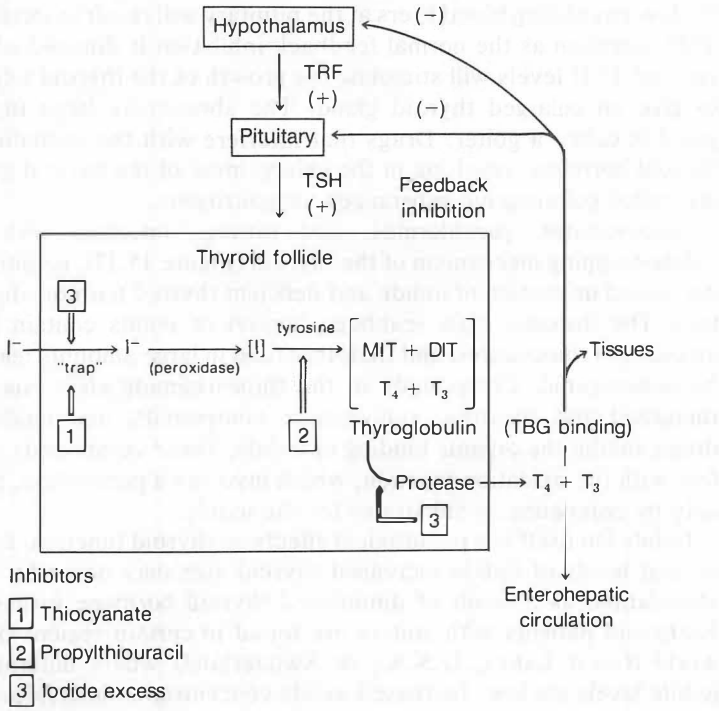


Figure 15.17
Regulation of thyroid hormone synthesis.

TSH through the activation of cAMP-mediated events. These reactions include the release of proteolytic enzymes from lysosomes, which hydrolyze thyroglobulin to yield the iodinated tyrosine and thyronine molecules. Approximately 100 μg of T_4 and T_3 in a ratio of 20 : 1 are released into the general circulation per day. A specific thyroid microsomal dehalogenase removes the iodide from the tyrosines but not from the thyronines and returns the halogen to the follicle for reutilization.

The uptake of iodide by the thyroid and the subsequent release of iodinated thyronines from thyroglobulin are both mediated through TSH stimulation. The conversion of T_4 to T_3 by thyroxine dehalogenase occurs in a number of tissues. This reaction is catalyzed by an enzyme that is different from the dehalogenase of thyroid tissue, since iodinated tyrosines are not substrates. In the liver T_4 and T_3 is further metabolized by deamination, decarboxylation, and conjugation (Figure 15.16). Both sulfate and glucuronide esters of the phenolic hydroxyl are formed in the liver. There is an extensive enterohepatic circulation of the conjugated hormone.

Inhibition of Thyroid Hormone Formation

When the formation of the thyroid hormones is blocked or inhibited, the low circulating blood titers at the pituitary will result in increased TSH secretion as the normal feedback inhibition is diminished. Increased TSH levels will stimulate the growth of the thyroid follicles to give an enlarged thyroid gland. The abnormally large thyroid gland is called a goiter. Drugs that interfere with the formation of thyroid hormone, resulting in the enlargement of the thyroid gland, are called goiterogenic substances, or goitrogens.

Thiocyanates, perchlorates, and nitrates interfere with the iodide-trapping mechanism of the thyroid (Figure 15.17), resulting in decreased utilization of iodide and deficient thyroid hormone formation. The *Brassica* class (cabbage family) of plants contain large amounts of thiocyantes, and their ingestion in large amounts tends to be goiterogenic. Compounds in the thiocarbamide class, such as thiouracil and thiourea, sulfonamide compounds, and imidazole drugs, inhibit the organic binding of iodide. These compounds interfere with the oxidation reaction, which involves a peroxidase, probably by competing as substrates for the iodide.

Iodide ion itself has paradoxical effects on thyroid function. Below normal levels of iodide increased thyroid size may occur by TSH stimulation, as a result of diminished thyroid hormone formation. Euthyroid patients with goiters are found in certain regions of the world (Great Lakes, U.S.A., or Switzerland) where natural soil iodide levels are low. Increased iodide concentration interferes with

TSH stimulatory effects on the thyroid tissue. Large doses of iodide are used to reduce thyroid size in certain hyperthyroid states (see Clin. Corr. 15.3).

Thyroid Hormone Function

All cells of the body, with the possible exception of adult brain and testes, are target cells for thyroid hormone. In most thyroid responsive cells T_3 binding sites have been demonstrated in the chromatin fraction of the nucleus. T_4 represents only 15% of the total iodothyronine bound to the nuclear receptor as the T_4 is converted in the cell to T_3 prior to binding. The receptor has not been well characterized. It has the high affinity ($K_d = 10^{-11}$ M) and low capacity for T_3, which would be expected for a receptor. In the normal (euthyroid) state only 15% of the total T_3 in the cell is bound, and ~50% of the available nuclear receptor sites are occupied. T_3 bound to the nuclear receptors is in equilibrium with the T_3 in the cytoplasm and circulating plasma. Unlike the steroid hormones, there appears to be no specific binding protein for T_3 in the cell cytoplasm. It is believed that increased mRNA production and increased protein synthesis are direct consequences of T_3-nuclear receptor binding, but evidence to support this concept is minimal and circumstantial. Another potential site for thyroid hormone action that has been proposed is the mitochondrion, based on the demonstration of high affinity binding. The mechanism of thyroid hormone action at the molecular level to explain the well-known physiological effects described below remains to be determined.

Thermogenesis and Oxygen Consumption

Increased heat production and oxygen consumption are characteristic for most tissues responding to thyroid hormone (brain, testis, and spleen excluded). Indeed, this feature was the basis for measurement [basal metabolic rate (BMR)] of thyroid hormone activity in the human prior to the development of more specific assay procedures for T_4 and T_3. The increased O_2 demand could be accounted for in part by the stimulation of membrane-bound sodium-potassium ATPase to maintain normal ionic gradients. Since thyroid hormone is present in cold-blooded as well as warm-blooded animals, its effect on thermogenesis requires more study to be fully integrated into present concepts of mode of action. Correlation of increased oxygen consumption with a mitochondrial site of thyroid hormone action has been considered, but the evidence that associates thyroid effects with energy production (electron transport system regulation) has been disputed. The ATPase enzyme activity in some obese patients have been found to be below normal. Since there is a direct correla-

CLIN. CORR. 15.3 USES OF RADIOACTIVE IODIDE

When a tracer dose of radioactive form of iodine (^{123}I), $T_{1/2}$ = 8 days) is administered orally, its rate of uptake by the thyroid gland ("iodide trap") is in response to the degree of stimulation by TSH. The rate of uptake is relatively high in the hyperfunctioning and low in the hypofunctioning gland (normal range is 5 to 50%). Similarly, the release of ^{123}I-thyronine from the thyroid to the blood can be monitored by a radioactivity scanning device 24 hours or more after the administration of a tracer dose of ^{123}I, since the release of thyroid hormone bound in thyroglobin stored in the gland is also stimulated by TSH.

Some types of thyroid cancer can be treated as an adjunct to surgery by the administration of larger doses of ^{131}I, since the radioactivity will concentrate in the thyroid tissue as a result of a very efficient trapping mechanism. The high energy of the β-emitter will destroy thyroid tissue selectively, particularly the more active cancer cells. Regions of high uptake are referred to as *hot spots* as distinct from inactive *cold spots*. It has been the practice in the past to restrict the use of radioactive iodine in young individuals, especially in women of childbearing age.

tion between thyroid hormone levels and ATPase activity, obesity in some patients has been suggested to be the result of a decreased energy and heat production due to diminished ATPase activity.

Metabolic Effects

Metabolic effects of T_4 and T_3 include alterations in carbohydrates, proteins, lipids, electrolytes, and water. Thyroid hormone effects on carbohydrate metabolism involve increased absorption of glucose balanced by increased glucose utilization. The net effect is one of hyperglycemia and an abnormal glucose tolerance curve. In protein metabolism a dual effect by thyroid hormone also is observed with anabolic activity during growth and development and catabolic activity (protein breakdown) in the hyperthyroid state. Cholesterol blood levels are high in hypothyroidism, and the high levels can be decreased with thyroid hormone administration. There is increased lipid utilization with thyroid hormone. Retention of water and electrolytes in the hypothyroid state can be reversed by thyroid administration.

Altering the thyroid hormone state in the human causes well-known changes in the central nervous system, in nerve and muscle function, in the gastrointestinal tract, and in the vascular system. The skin is a good indicator of the thyroid state. In hyperthyroidism the skin is smooth, warm, and moist as a result of vasodilation. In contrast, the skin is cold and has a rough texture due to vasoconstriction in the hypothyroid state. The characteristic accumulation of fluid and mucopolysaccharides, with the resulting puffiness (pitting edema) of the skin, gives rise to the adult hypothyroid state, myxedema.

The heart reflects the changes in thyroid state, having a slow rate and decreased blood flow in the hypothyroid condition. A major toxic effect of thyrotoxicosis is on the heart, with increased heart rate, cardiac hypertrophy, and cardiac contractility.

The thyroid hormones act in conjuction with pituitary growth hormone as the principal anabolic agents during growth and in maintaining protein stores. Synergistic effects of the two can be demonstrated on protein synthesis in the liver and on transamidinase (guanidine transferase) activity in the kidney. The important role of thyroid hormone in human development is apparent in *cretinism,* a condition brought about by thyroid deficiency during the prenatal period, resulting in serious detriment to both mental and physical development in the growing child.

BIBLIOGRAPHY

Receptor Models

Baxter, J. D., and MacLoed, K. M. Molecular basis for hormone action, in P. K. Bondy and I. N. Rosenberg (eds.), *Metabolic control and disease*. Philadelphia: W. B. Saunders, 1980, p. 140.

Clark, J. H., and Peck, R. J., Jr. Steroid hormone receptors: basic principles and measurement, in B. W. O'Malley and L. Birnbaumer (eds.), *Receptors and hormone action*. New York: Academic, 1977, p. 383.

Jensen, E. V., and Jacobson, H. I. Basic guides to the mechanism of estrogen action. *Rec. Prog. Horm. Res.* 18:387, 1962.

Schrader, W. T. Methods for extraction and quantification of receptors, in B. W. O'Malley and J. G. Hardeman (eds.), *Methods Enzymol.* 36:187, 1975.

Thyroid Hormones

Edelman, L. S. Thyroid thermiogenesis. *N. Engl. J. Med.* 290:1303, 1974.

Oppenheimer, J. H., Schwartz, H. L., Surks, M. I., Koerner, D. H., and Dillman, W. H. Nuclear receptors and initiation of thyroid hormone action. *Rec. Prog. Horm. Res.* 32:529, 1976.

Steroid Hormones

British Medical Bulletin. *Radioimmunoassay and Saturation Analysis,* vol. 1, 1974.

Dorfman, R. I., and Ungar, F. *Metabolism of steroid hormones.* New York: Academic, 1965.

Goto, J., and Fishman, J. Participation of a nonenzymatic transformation in the biosynthesis of estrogens from androgens. *Science* 195:80, 1977.

Jensen, E. V., and DeSombre, E. R. Estrogen-receptor interaction. *Science* 182:126, 1973.

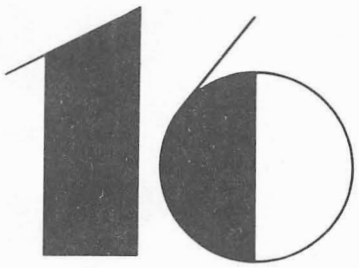
Leavitt, W. W., and Clark, J. H. (eds.), *Steroid hormone receptor systems.* New York: Plenum, 1979.

O'Malley, B. W. Mechanisms of action of steroid hormones. *N. Engl. J. Med.* 284:370, 1971.

Ortiz de Montillano, P. R., and Kunze, K. L. Self catalyzed inactivation of hepatic cytochrome P_{450} by ethynyl substrates. *J. Biol. Chem.* 255:5578, 1980.

Sheridan, P. J. The nucleus interstitialis striae terminalis and the nucleus amygdaloideus medialis: prime targets for androgen in the rat forebrain. *Endocrinology* 104:130, 1979.

Wilson, J. D. Recent studies on the mechanism of action of testosterone. *N. Engl. J. Med.* 287:1284, 1972.



Biochemistry of Hormones II

FRANK UNGAR

PEPTIDE HORMONES

16.1 PLASMA MEMBRANE RECEPTORS

Two principal types of hormone–receptor interactions were introduced in Chapter 15. In one type, lipid-soluble substances, including steroid hormones, 1,25-dihydroxycholecalciferol, and thyroid hormones, pass freely through the plasma membrane and bind to a cytoplasmic receptor. The hormone–receptor complex translocates to the chromatin of the cell nucleus to initiate the DNA-directed synthesis of mRNA. The second type of hormone–receptor interaction involves water-soluble substances, which are not permeable to the cell plasma membrane; their first contact with the cell is to bind to protein receptors on the outer membrane surface.

In order for the message to reach intracellular sites, the hormone–receptor complex could be internalized (endocytosis), or the information could be transferred at the membrane site (transduction) and carried by other mediators or messengers within the cell compartments. Substances that are considered to participate in the intracellular transfer of information include the calcium ion (Ca^{2+}), the prostaglandins, the cyclic nucleotides, and the proteins with

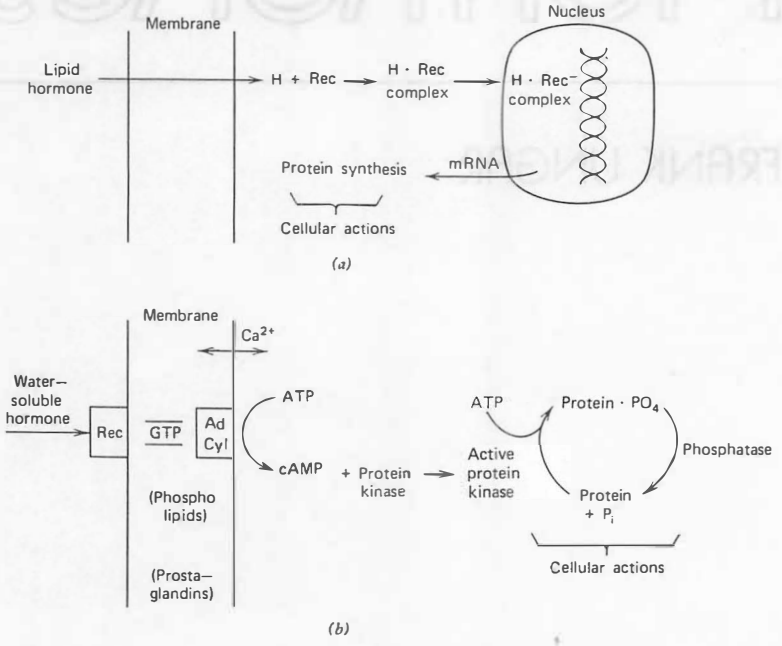


Figure 16.1
Initial interactions of hormones with target cells.
(a) Lipid hormones bind to receptors within the cell and as a hormone receptor complex initiate DNA-directed synthesis of mRNA (examples: steroid hormones, vitamin D, thyroid hormones). (b) Water-soluble hormones bind to specific receptors on the cell plasma membrane. By activation of adenylate cyclase and protein kinase, a number of cellular processes are stimulated (examples: glucogen, epinephrine, ACTH). See text for actions of Ca^{2+}, Mg^{2+}, GTP, phospholipids, and prostaglandins.

which they interact. Examples of the participation of these mediators in the expression of hormone action for the water-soluble hormones are presented in this chapter. (See Figure 16.1.)

The plasma membrane presents a barrier to the entrance of water-soluble molecules, and within the cell the subcellular membranes restrain free access of substances. The membrane, by enfolding materials from the exterior, can internalize them (endocytosis), or by coalescing with vesicles from within, can transport materials out (exocytosis) of the cell. Cellular activity is enhanced by removing the membrane restraint by increasing its permeability to the flow of ions or organic substances. This is described in terms of creating membrane pores, channels, or gaps by which water-soluble substances can be transferred into the cell or between the cell compartments created by membranes. (See Figure 16.2.) Hormone binding to the cell membrane or by indirectly affecting subcellular membrane permeability is in effect removing a restraint to cell activity or subparticle interaction. Water-soluble hormones, neurotransmitters, or other agents (lectins) bind to a cell surface receptor as a result of a noncovalent association with a relatively high affinity to initiate a set of membrane events. As a result, there is a release of mediators (secondary messengers) that react with cytoplasmic components. Most polypeptide hormones stimulate the production of cAMP as

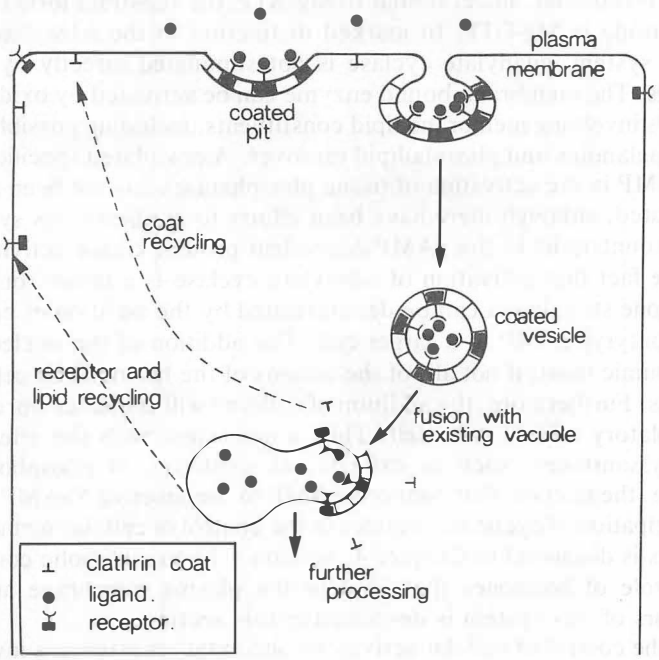


Figure 16.2
Internalization of substances through the cell membrane using coated pits and coated vesicles.
Pathway by which cholesterol-lipoprotein complex and peptide hormones can enter the cell. Inside the cell the ingested materials, receptors, hormones, and cholesterol, are processed by lysosomal hydrolytic enzymes. Processed molecules can then be transferred to other cell compartments.
From B. Pearse, *Trends Biochem. Sci.*, 5:26, 1980. Reproduced with permission.

the secondary messenger. There are important exceptions to this, however. Insulin and prolactin have been shown to bind to plasma receptors without a direct stimulation of adenylate cyclase. Their membrane interactions are described in a later section.

Both receptor and adenylate cyclase can be isolated and purified as separate proteins in a number of different cell membranes. Cells are lysed under hypoosmotic conditions, the cell contents are removed, and the heavy tissue fragments, which sediment by centrifugation, partially contain intact membrane particles (ghosts). Liver cell ghosts or fat cell ghosts are commonly used as model systems. Since the hormone receptor and adenylate cyclase are present and active in the same particle, it is possible to measure both hormone binding to the receptor as well as cAMP production, as a membrane response to hormone stimulation. An unusual feature of adenylate cyclase reaction, which forms cAMP and PP_i, is the stimulation of the enzyme by halides, particularly the fluoride ion. Maximal stimulation is achieved in the presence of F^-, which is not increased further by hormone stimulation. The organization of the enzyme complex in the membrane is important, since the enzyme in a soluble free form still retains activity, but no longer can be stimulated by hormones.

Guanylate cyclase activity in the conversion of GTP to cGMP and PP_i can be maximally stimulated in the presence of added Mn^{2+} ion. Mg^{2+} is required, since, similar to Mg-ATP, the substrate form of the nucleotide is Mg-GTP. In marked distinction to the adenylate cyclase system, guanylate cyclase is not stimulated directly by hormones. The membrane-bound enzyme can be activated by oxidative events involving membrane lipid constituents, including possibly the prostaglandins and phospholipid turnover. A postulated specific role of cGMP in the activation of tissue phosphatases has not been demonstrated, although there have been efforts to implicate this system as a counterpart to the cAMP-dependent protein kinase activity.

The fact that activation of adenylate cyclase is a major route of hormone stimulation can be demonstrated by the addition of cAMP or dibutyryl cAMP to a target cell. The addition of the nucleotide will mimic most, if not all, of the actions of the hormone on cellular events. Furthermore, the addition of caffeine will enchance hormone stimulatory effects on a cell. This is consistent with the effect of methylxanthines, such as caffeine, as inhibitors of phosphodiesterase, the enzyme that converts cAMP to the inactive 5′-AMP. The participation of cyclic nucleotides in the control of cellular metabolic events is discussed in Chapter 4, Section 4.10 on metabolic control. The role of hormones that bind to the plasma membrane as activators of this system is described in this section.

In the control of cellular activity by secondary messengers involv-

ing the same cyclic nucleotides, Ca^{2+} ion, Mg^{2+} ion, and prostaglandins, the question arises as to the specificity of cell response and specialization of cellular activity. Assuming the genetic makeup of each cell is identical and it possesses a common potential, the expression of cellular activity is the result of the pattern and amounts of enzyme activities that are free to be expressed. This probably occurs early in development by removal of restraints at the level of genome, which will vary among the different cell types.

If the pattern of response of the cell is formed early in the differentiation of the organism, at the other end of the spectrum the specificity of response is determined by the presence of cell receptors at the plasma membrane, which initiates those events leading to that cell response. The ability of a hormone to stimulate a cell is regulated by the type and amount of receptor present on the cell membrane. It is probable that each cell has the potential to respond to any hormone if a sufficient number of receptors are present and active on the membrane surface.

The number of receptors on the cell membrane is relatively low, on the order of 5,000 to 10,000 per cell. In most systems relatively few receptors, 10–20% of the total present, need to be occupied to produce a maximum response of the cell. The presence of "spare receptors" is considered a feature which ensures by the redundancy of the system that a given cell will respond to a hormone stimulus. The number of receptors per cell is not fixed and can be shown to vary under different physiological conditions as determined by the use of Scatchard plots.

The terms desensitization and down regulation have been applied to a decrease in cell response to hormone stimulation. Down regulation is observed whenever cells are exposed under chronic conditions to relatively high tissue concentrations of hormone. Explanations for the occurrence of down regulation are not well defined and may vary; for example, (1) increased number of receptor sites occupied by hormone, (2) the presence of antibodies that may block the receptor site, (3) the turnover of receptor sites from the membrane surface to the membrane interior, and (4) the decreased synthesis and turnover of receptor protein by the cell. At present there is insufficient evidence to support any of these postulates.

The concentration of a membrane receptor that is saturated by hormone levels at 10^{-8} to 10^{-10} M must be correspondingly low and would require a degree of purification not yet possible to achieve by present methods. Using affinity chromatography with the hormone attached as a ligand to an inert phase, the best receptor preparations achieved from mammalian tissues are, with a few exceptions, still only 10–50% pure. Model systems for the study of hormone receptors and their activation of membrane-bound adenylate cyclase have

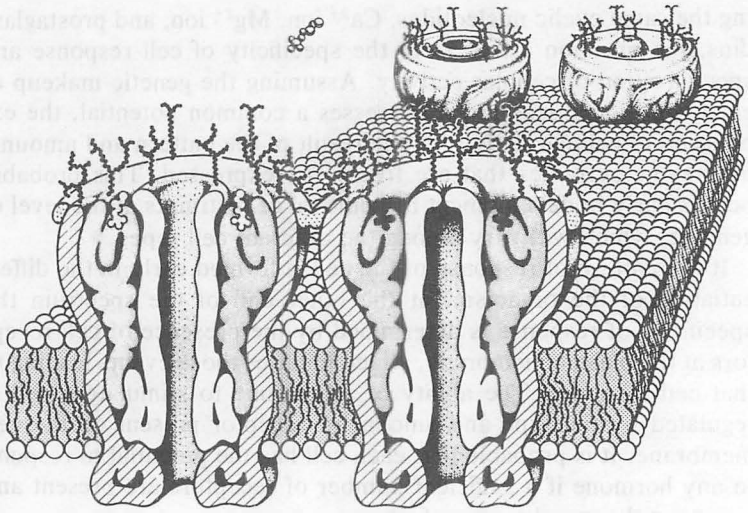


Figure 16.3
A depiction of the transmembrane pentameric receptor for acetylcholine.
Channel open (*left*), *channel closed* (*right*). From J. Lindstrom, *Advan. Immunol.*, 27:1, 1979. Reproduced with permission.

been limited in the past, yet valuable information has been obtained from other membrane receptor systems, even though they may not be directly applicable to mammalian hormone systems. The receptor system for acetylcholine and the stimulation of adenylate cyclase by cholera toxin are useful illustrations.

Acetylcholine Receptor

The acetylcholine receptor (Figure 16.3) has been purified to the greatest degree of homogeneity and its subunit character fully determined. The unusually high concentration of receptor in the electric organ of the sting ray (*Torpedo*) and eel (*Electrophorus electricus*) allowed purification of the acetylcholine receptor to homogeneity after a 500-fold purification. Due to their very low concentrations in cells, mammalian hormone receptors will require thousandsfold purification to achieve the same degree of homogeneity.

The acetylcholine receptor is a pentamer of four homologous proteins of mol wt 255,000, with subunits of 2 × 40K, 50K, 60K, and 65K, forming a structure of $\alpha_2\beta,\gamma,\delta$. The receptor is a transmembrane protein with a carbohydrate portion on the outer (synaptic) face of the membrane and forms a channel for ion transfer. Acetylcholine is a neurotransmitter released at nerve terminals in response to an electrical signal at a nerve–muscle structure known as a motor end plate. Specific receptors are present across a synaptic gap in the postsynaptic membrane, which receives acetylcholine, binds and hydrolyzes it by reaction with an esterase. This interaction opens channels in the receptor to allow Na^+ and K^+ to pass through.

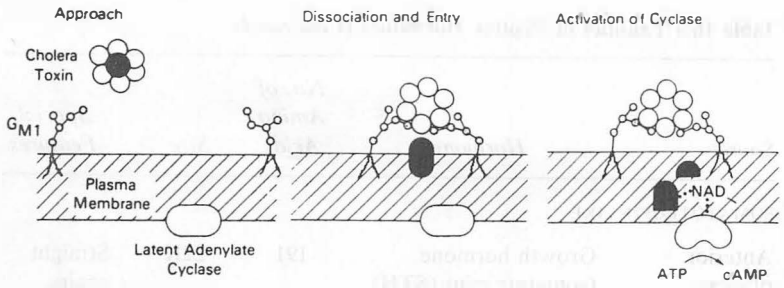


Figure 16.4

Action of cholera toxin on adenylate cyclase. *Toxin-B subunit binds to G_{M1} receptor, releasing A subunit, which enters membrane, and in presence of NAD causes ADP-ribosylation of subunit of adenylate cyclase system. This occurs at the GTP binding site and inhibits GTPase activity. The result is an irreversible activation of adenylate cyclase.*

From P. H. Fishman and R. O'Brady, *Science*, 194:906, 1976. Copyright 1976 by the American Association for the Advancement of Science.

Cholera Toxin

Cholera toxin is a protein (84K) secreted by *Vibrio cholerae* and is responsible for the clinical effects of cholera in the intestines. It has a structure of three peptides $\alpha\gamma\beta n$, arranged into an A (α, γ) subunit and a B (βn) subunit. (See Figure 16.4.)

The binding site of the molecule on the β subunit binds to a GM_1 ganglioside of the outer membrane. The A subunit of the molecule enters the membrane and splits into A_1 and A_2 subunits. The A_2 subunit binds irreversibly to the GTP binding protein (Chapter 4) of adenylate cyclase in the inner membrane. This results in an inhibition of the GTPase activity, which normally would turn off adenylate cyclase. The irreversible binding of the A_2 fragment results in the extreme stimulation of adenylate cyclase to produce constant formation of cAMP. This alters the permeability of the intestinal membrane and allows copious amounts of water and electrolytes to be lost, which accounts for the typical severity of the disease.

Table 16.1 Families of Peptide Hormones

Source	*Hormone*	*No. of Amino Acids*	*Size*	*Special Features*
GLUCOSE REGULATION				
Pancreas	(β cell) Insulin	51	6K	3 Disulfide bridges
	(α cell) Glucagon	29	35K	Straight chain
HYPOTHALAMUS–PITUITARY				
Posterior pituitary	Oxytocin	9	1K	1 Disulfide bridge
	Vasopressin	9	1K	1 Disulfide bridge

CLIN. CORR. 16.1 MEMBRANE AND NUCLEAR SITES OF HORMONE ACTION

The binding of a peptide hormone to a plasma membrane receptor or that of steroid and thyroid hormones to a nuclear receptor does not portend adequately the full range of hormone action normally encountered in clinical medicine. Long-terms effects, particularly those associated with regulation of cell division and tissue growth, have not been studied extensively and are poorly understood. Subsequent to membrane receptor binding, peptide hormones can be internalized by a process referred to as *receptor mediated endocytosis*, a process of membrane enfolding to form cytoplasmic vesicles, which transports the hormone receptor to the lysosome and other possible subcellular compartments. Plasma membrane receptors have a complex subunit structure that may involve other transport functions in addition to hormone binding. The actions of insulin, prolactin, and other hormones that have been described at the intracellular level may be a consequence of additional events other than those associated with cyclic AMP-mediated protein kinase-phosphatase interactions. Similarly, membrane associated events can be associated with administration of steroid and thyroid hormones, although the mechanisms involved cannot be described at the level of molecular interaction.

Table 16.1 Families of Peptide Hormones (*Continued*)

Source	*Hormone*	*No. of Amino Acids*	*Size*	*Special Features*
SOMATOTROPIN/LIKE				
Anterior pituitary	Growth hormone (somatotropin) (STH)	191	22K	Straight chain; 2 disulfide bridges
	Prolactin	198	22K	3 Disulfide bridges
Placenta	Placental lactogen (somatomammotropin)	191	22K	2 Disulfide bridges
ACTH/ENDORPHIN				
Anterior pituitary (hypothalamus)	Adrenocorticotropin (ACTH)	39	4.5K	Straight chain
	Melanocyte stimulating (α-MSH)	13	1.8K	Acetyl on Ser^1
	β-Lipotropin	91	9.5K	Straight chain
(hypothalamus)	β-Endorphin	30	4.0K	Straight chain
GLYCOPROTEIN HORMONES				
Anterior pituitary	Thyroid stimulating (TSH)	209	27K	CHO 15–30%
	Follicle stimulating (FSH)	236	32K	α subunit
	Luteinizing (LH)	215	30K	β subunit
Placenta	Human chorionic gonadotropin (HCG)	231	46K	α, 92 aa β, 139 aa
CALCIUM REGULATION				
Parathyroid	Parathyroid hormone (PTH)	84	9.5K	Straight chain
Parathyroid and thyroid	Calcitonin	32	3.6K	Straight chain

The binding subunit B has some amino acid homology to the β subunits of the pituitary glycoprotein hormones, TSH, HCG, FSH, and LH. (See Table 16.1.) However, this homology also extends to the serine proteases, which have no membrane function, so that amino acid homology alone may not be significant. There is no analogous fate of the α subunits of the glycoprotein hormones comparable to that of the A subunit of cholera toxin. In the A subunit interaction, NAD is converted to form a covalently bound ADP-ribosylated protein with the release of nicotinamide, resulting in the irreversible stimulation of adenylate cyclase. No such reaction occurs with the subunit peptides of the pituitary hormones. Although the cholera toxin will stimulate hormone-responsive tissues and provides important insight into the mode of action of adenylate cyclase, it does not reflect on the manner of interaction of the pituitary hormones with their respective receptors or their subsequent mode of stimulation of adenylate cyclase. (See Clin. Corr. 16.1.)

16.2 HORMONE REGULATION OF CALCIUM

The intracellular concentration of calcium at 10^{-6} to 10^{-7} M is maintained by an active Ca^{2+}/Mg^{2+} ATPase system, which extrudes Ca^{2+} into an extracellular environment where the (Ca^{2+}) ion is 10^{-3} M. The dominant role exerted by calcium and calmodulin within the cell with respect to the regulation of a number of enzyme systems, particularly with respect to hormone action on cyclic nucleotide regulation, has been reviewed in Chapter 4, Section 4.10. Extracellular calcium is rigidly controlled as well, and this discussion describes those hormonal factors that are involved in this regulation. The plasma (Ca^{2+}) level is maintained closely at 10 mg/dl by three substances: (1) parathyroid hormone (PTH), (2) calcitonin, and (3) cholecalciferol or vitamin D. The actions of these three agents to achieve homeostasis are exerted in specific tissues of the body in a concerted fashion, with the ultimate control maintained by the levels of blood calcium itself.

The major physiological counterion to calcium is phosphate. The solubility product of calcium phosphate is exceedingly low, and a large increase in one ion will cause a reciprocal decrease in the other; otherwise, the insoluble calcium phosphate will precipitate.

At blood pH 7.4 the most common form of phosphate is $[HPO_4^{2-}]$. At an acid pH (pH ~6.0) of urine $[H_2PO_4^-]$ is the common form excreted. The insoluble $(Ca^{2+})_3(PO_4^{3-})_2$ is found in bone and teeth in a crystalline lattice form of hydroxyapatite, $Ca_{10}(PO_4)_6(OH)_2$, as a mixture that also contains the elements Cl^-, F^-, CO_3^{2-}, Na^+, K^+, and Mg^{2+}. In the blood, one-half of the calcium is bound to plasma proteins. It is the remaining 50% of free $[Ca^{2+}]$ ion that is physiologically effective. Bone contains over 99% of the body calcium and is therefore the main reservoir to be used when calcium is required. Dietary calcium is usually more than sufficient to maintain normal blood $[Ca^{2+}]$ levels. The amount of $[Ca^{2+}]$ in foods, however, is not always readily available; it may be present in an insoluble form, or it may not be absorbed by the intestinal mucosa. Normally $[Ca^{2+}]$ is excreted in the feces. The urinary excretion of $[Ca^{2+}]$ is usually minimal but increases when blood levels of $[Ca^{2+}]$ are raised.

Parathyroid Hormone and Calcitonin

The blood $[Ca^{2+}]$ level determines the rate of secretion of parathyroid hormones (PTH). Low levels of blood $[Ca^{2+}]$ release the hormone contained in vesicles in the parathyroid cell. The hormone is a straight-chain polypeptide (mol wt 9,500) containing 84 amino acids. It is formed as a preprohormone and is cleaved to a prohormone (90 amino acids) and to its final active form (84 amino acids), as it is transferred from the endoplasmic reticulum to Golgi apparatus to secretory vesicle. Once secreted into the blood, PTH has a relatively short half-life, measured in minutes, as is the case for most peptide hormones. Inactive fragments, which can be detected by radioimmunoassay using an antibody of the entire PTH (1–84), are found in target cell membranes or circulating in the plasma. The major portion contains the COOH end of the molecule and is biologically inactive. The smaller NH_2 end of the molecule retains some biological activity. The correspondence of biological activity of PTH to the amount of hormone as determined by radioimmunoassay may be quite variable for this reason. (See Figure 16.5.)

Calcitonin is a peptide (mol wt 3,600) containing 32 amino acids in a straight chain, which is produced from a larger peptide in a prohormone form. Calcitonin is secreted from cells of the parathyroid and the thyroid gland in response to blood containing high levels of $[Ca^{2+}]$. The release of calcitonin leads to a gradual lowering of blood $[Ca^{2+}]$ levels. This action of calcitonin is the result of decreased removal (resorption) of calcium from bone. A primary action of PTH is on the calcification system of bone with a direct effect on osteolytic activity leading to bone $[Ca^{2+}]$ resorption. Both hormone actions involve cAMP in an undetermined manner in which PTH increases,

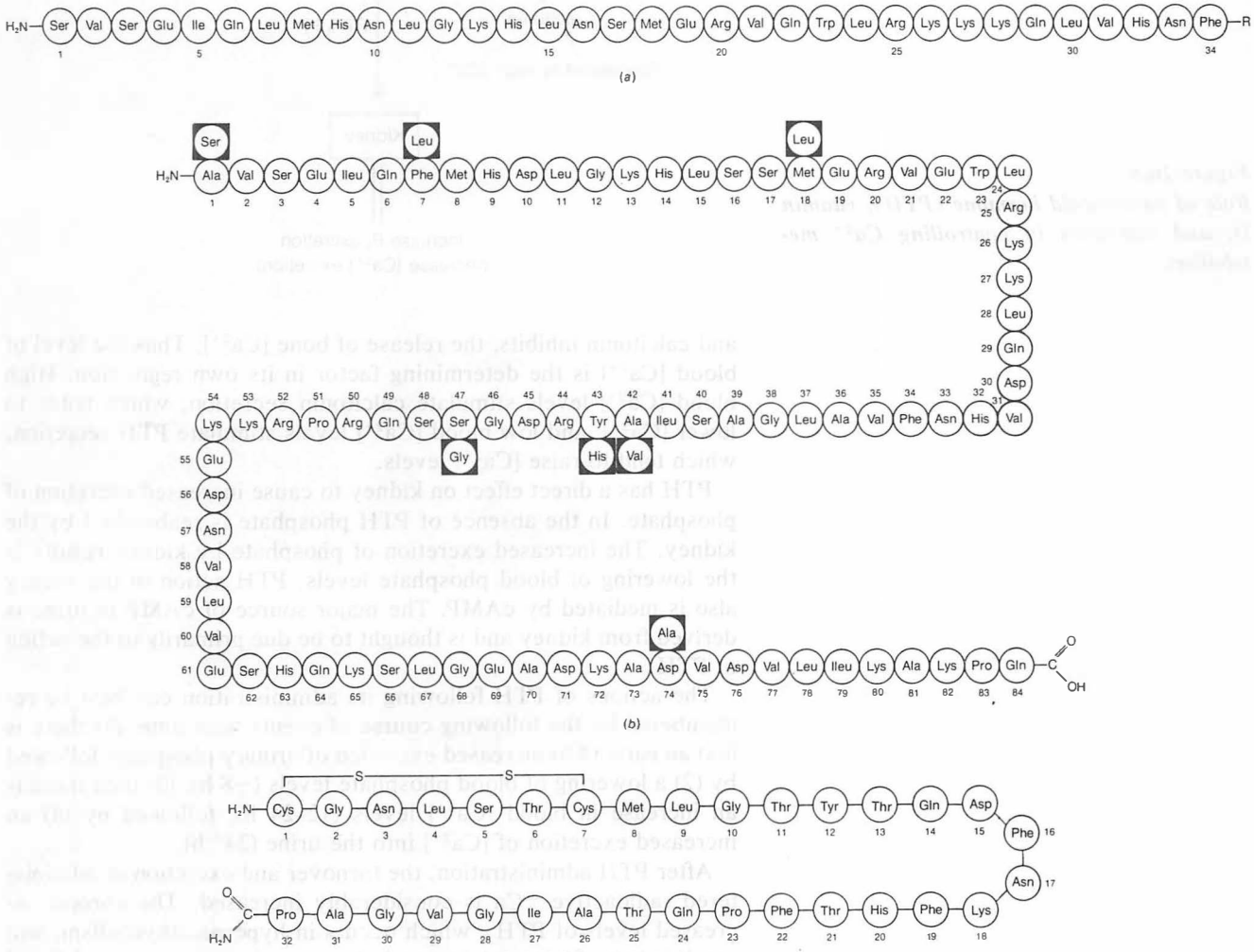


Figure 16.5
(a) Amino acid sequence (1–34) of human parathyroid hormone compared to that of (b) bovine (white circles) and ovine (dark circles) hormones (1–84). (c) Amino acid sequence of human thyrocalcitonin.

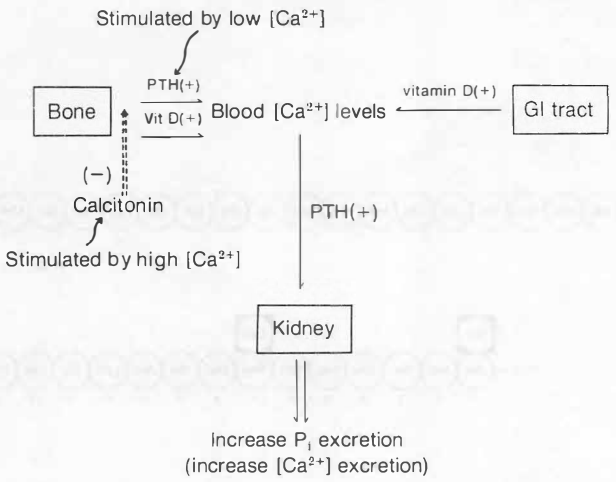


Figure 16.6
Role of parathyroid hormone (PTH), vitamin D, and calcitonin in controlling Ca^{2+} metabolism.

and calcitonin inhibits, the release of bone $[Ca^{2+}]$. Thus the level of blood $[Ca^{2+}]$ is the determining factor in its own regulation. High blood $[Ca^{2+}]$ levels stimulate calcitonin secretion, which tends to lower $[Ca^{2+}]$, and low blood $[Ca^{2+}]$ levels stimulate PTH secretion, which tend to raise $[Ca^{2+}]$ levels.

PTH has a direct effect on kidney to cause increased excretion of phosphate. In the absence of PTH phosphate is reabsorbed by the kidney. The increased excretion of phosphate by kidney results in the lowering of blood phosphate levels. PTH action in the kidney also is mediated by cAMP. The major source of cAMP in urine is derived from kidney and is thought to be due primarily to the action of PTH.

The actions of PTH following its administration can best be remembered by the following course of events with time: (1) there is first an early (4 h) increased excretion of urinary phosphate followed by (2) a lowering of blood phosphate levels (~8 h); (3) then there is an increase in blood $[Ca^{2+}]$ levels (12–24 h), followed by (4) an increased excretion of $[Ca^{2+}]$ into the urine (24+ h).

After PTH administration, the turnover and excretion of administered radioactive ^{35}Ca is considerably increased. The chronic increased levels of PTH , which occurs in hyperparathyroidism, will result in loss of bone calcium and often can result in renal calculi and in the deposition of calcium phosphate in the soft tissues. Hypoparathyroidism can occur by the removal of the parathyroid glands during thyroidectomy. In the absence of PTH blood $[Ca^{2+}]$ levels can fall to a level of 7 mg/dl, which may precipitate a tetany, a rapid uncontrolled contraction of skeletal muscle due to the effect of low $[Ca^{2+}]$ levels on neuromuscular irritability. (See Figure 16.6.)

Vitamin D

Parathyroid hormone activity, which would be increased at lower blood [Ca^{2+}] levels, exerts an effect in the kidney to stimulate 1α-hydroxylation of a vitamin D precursor to produce 1,25-dihydroxycholecalciferol. This, the most active form of vitamin D, acts on the mucosal cells to increase the intestinal transport of [Ca^{2+}]. Vitamin D (or cholecalciferol) is produced in the skin as the result of irradiation of the precursor, 7-dehydrocholesterol. Cholecalciferol (Figure 16.7) is transferred to the liver, where 25-

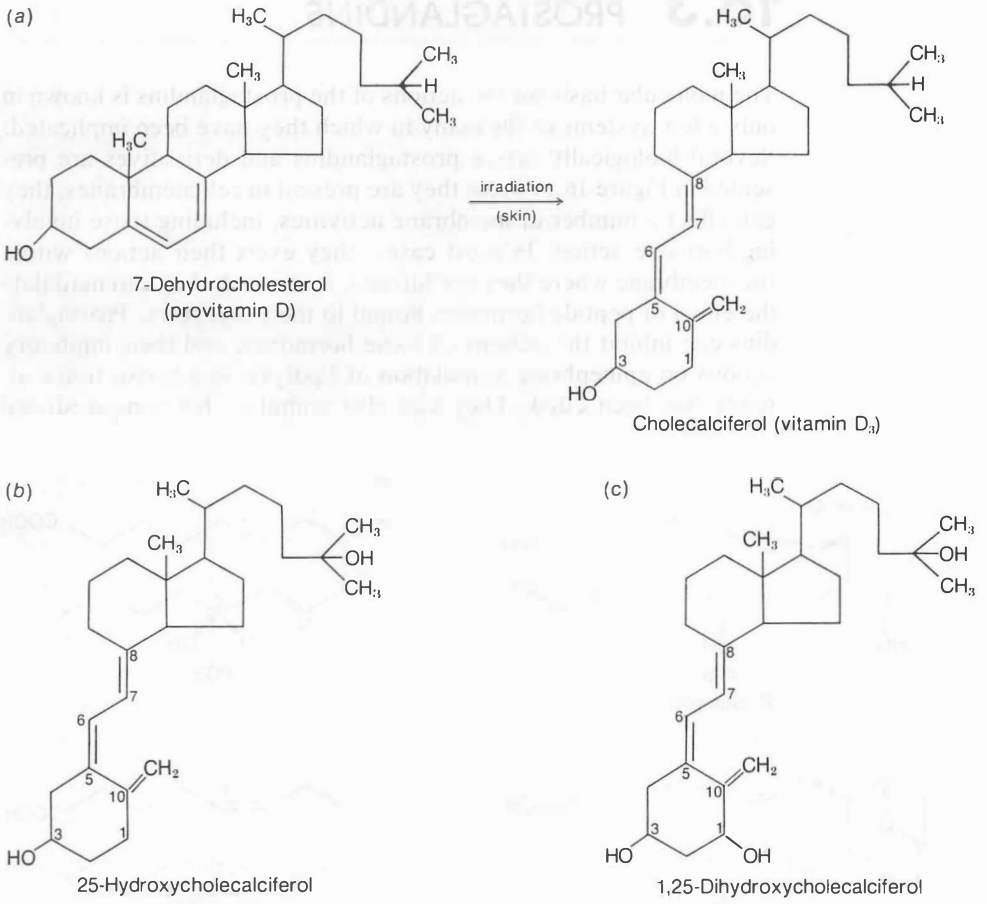


Figure 16.7
(a) Activation of cholecalciferol (vitamin D_3); (b) hydroxylation occurs in the liver to form 25-hydroxycholecalciferol; and (c) in the kidney to form 1,25-dihydroxycholecalciferol.

hydroxycholecalciferol is formed, which in turn is converted in the kidney by 1α-hydroxylation to the 1,25-dihydroxy form of vitamin D. It is thought that the action of vitamin D on the mucosal cell results in the formation of a calcium binding protein, which enhances the absorption of Ca^{2+} by the gut. 1,25-Dihydroxycholecalciferol also is active on bone and enhances the mobilization of calcium to the blood.

16.3 PROSTAGLANDINS

The molecular basis for the actions of the prostaglandins is known in only a few systems of the many in which they have been implicated. Several biologically active prostaglandins and derivatives are presented in Figure 16.8. Since they are present in cell membranes, they can affect a number of membrane activities, including those involving hormone action. In most cases, they exert their actions within the membrane where they are formed. As a result they can modulate the effect of peptide hormones bound to their receptors. Prostaglandins can inhibit the actions of some hormones, and their inhibitory actions on epinephrine stimulation of lipolysis in adipose tissue already has been cited. They can also stimulate hormone-mediated

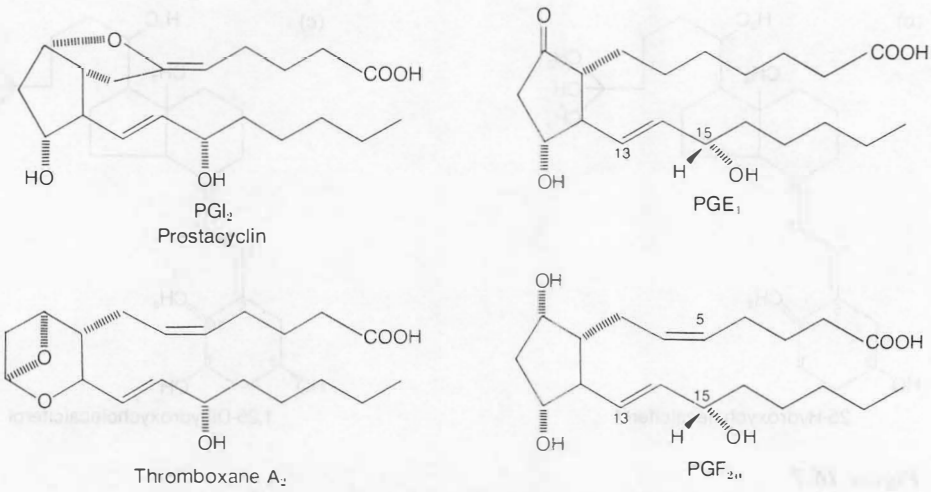


Figure 16.8
Biologically active prostaglandins.

events; a slight ACTH-like effect on adrenal steroidogenesis has been observed. Some of these actions are related to the stimulation of adenylate cyclase, and others, as with epinephrine, to the inhibition of adenylate cyclase. Their exact role as modulators of hormone action with respect to membrane function and activity is yet to be determined.

Prostaglandins are formed from polyunsaturated fatty acids released from phospholipids in the plasma membrane of cells. The history of the discovery and elucidation of the structure of prostaglandins parallels events during the last half-century, a period in which the role of polyunsaturated fatty acids was proposed and determined to be essential for normal human nutrition. Prostaglandin activity was initially discovered in human semen using a bioassay procedure that measured smooth muscle contraction. The details of the synthesis of the prostaglandins are presented in Chapter 10, Section 10.5.

Regulation of Prostaglandin Synthesis

Membrane phospholipids such as phosphatidylcholine, and particularly phosphatidylinositol, provide the precursors to the prostaglandins. Release of the free fatty acids C_{20}: 3, C_{20}: 4, and C_{20}: 5 from the 2-carbon of the phospholipid is the result of activity by phospholipases. Ca^{2+} ions are required for enzyme activation. One model proposed for platelet aggregation involves the influx of Ca^{2+}, which binds to the calcium binding protein, calmodulin, to stimulate phospholipase. This stimulation could involve a cAMP-mediated event. Phospholipase A_2 hydrolyzes phosphatidylinositol with the release of arachidonic acid. The conversion of arachidonic acid to thromboxane by blood platelets leads to aggregation. Platelet aggregation is counteracted by the formation of prostacyclin in the blood vessel wall and this prevents platelets sticking to the lining. In blood vessel injury, the formation of leukotriene would inhibit prostacyclin, and the injury or vessel "hole" could then be walled off.

Both oxidative and reductive processes at the membrane site have important consequences for prostaglandin metabolism and inactivation (Figure 16.9). The peroxidative enzymes cyclooxygenase and lipoxygenase in conjunction with reductive reactions with glutathione or the —SH group in cysteine of proteins form a series of short-lived intermediates with significantly different biological activities. The physiological signals at the membrane site to initiate these events have not been determined. Ascorbic acid, nitrates, and other oxygen-containing substances can stimulate prostaglandin synthesis. (See Figure 16.10.)

Inhibitors of prostaglandin synthesis are known as well. Cortisol

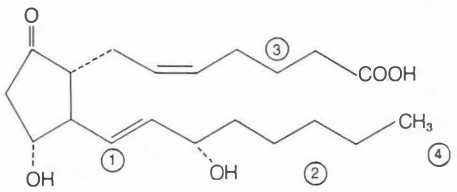
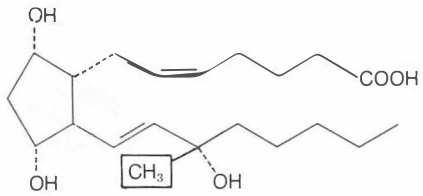


Figure 16.9
Metabolism of the prostaglandins leading to inactivation.
1, Reduction of Δ^{13} double bond; 2, Oxidation of C_{15}-OH to the ketone; 3, β-oxidative cleavage of side chain; 4, Oxidation of —CH_3 side chain.



15 (S) 15-Methyl-$PGF_{2\alpha}$

Figure 16.10
Synthetic prostaglandin.
Methyl group protects oxidation of 15-hydroxyl and enhances the biological activity.

inhibits phospholipase A_2 activity to inhibit the release of arachidonic acid (Figure 16.11). The reaction between arachidonic acid and endoperoxide formation catalyzed by cyclooxygenase is inhibited by the drugs aspirin and indomethacin. Aspirin is known to acetylate the cyclooxygenase of platelets and to inhibit thromboxane A formation. The inhibitory activities by these substances in vitro on prostaglandin synthesis have been correlated with some of their known physiological effects in vivo (Figure 16.11).

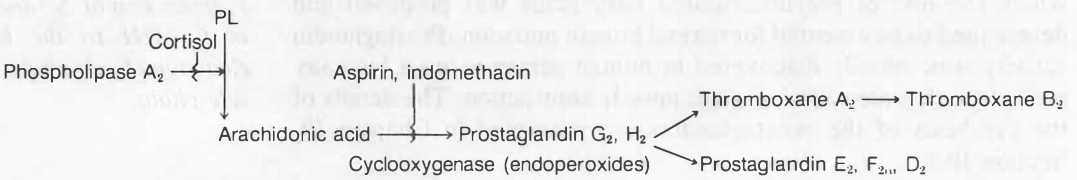


Figure 16.11
Sites of inhibition by cortisol and by aspirin or indomethacin in the biosynthesis of prostaglandins and thromboxanes.

Biological Actions

Different effects of the PGE and PGF analogs are observed on smooth muscle and membranes, which results in dilation or constriction of blood vessels in a large number of tissues, including uterus, lung, and kidney. Two early significant observations provided the stimulus for further extensive research. The first applications, which generated worldwide interest, were reported by an obstetrician in Uganda who published a paper on the use of prostaglandins for uterine contraction in aiding childbirth, and for the induction of abortions in early pregnancy. The second report demonstrated that prostaglandins in vitro inhibited the effect of epinephrine on the release of free fatty acids from fat cells. This observation placed the role of prostaglandins at a membrane site, relating its activity to that of the membrane-bound adenylate cyclase system. It was suggested that a coupler, or transducing system, which linked hormone binding to its receptor to the stimulation adenylate cyclase was related to phospholipid turnover or to the prostaglandins derived from the phospholipids.

The many diverse actions of the prostaglandins at the site of the cell membrane on smooth muscle and on blood vessels will vary according to whether the compounds are of the PGE or PGF series. PGE will stimulate cAMP production and therefore a number of

reactions initiated by cAMP. $PGF_{2\alpha}$ stimulates contraction of the endometrium of the uterus and is a luteolytic agent; it is also a bronchial dilator in lung tissue. At the present time it is not known whether these and many other effects attributed to PGE and PGF compounds represent true physiological events, or whether these substances are merely stable end products of a series of short-lived reactants involving unstable forms of oxygen-containing intermediates such as endoperoxides, thromboxanes, and prostacyclins. The leukotrienes represent still another class of short-lived intermediates, which are linked to highly sensitive tissue substituents.

Their ubiquity, their short duration of activity, and the physiological importance of the systems involved, hold great promise for prostaglandins and their analogs, both as important links in the study of the complex nature of tissue and metabolic regulation and as models for the synthesis of new classes of biologically active therapeutic agents.

16.4 PANCREATIC HORMONES

The levels of blood glucose are maintained relatively constant by the actions of liver to take up or release glucose. In the liver glucose is converted to glucose 6-phosphate, a key intermediate leading to a number of different metabolic pathways: (1) formation of glycogen, (2) metabolism via glycolysis, (3) oxidation by way of the glucose shunt, or (4) released to the blood as free glucose by action of glucose 6-phosphatase. The fate of the glucose 6-phosphate is determined essentially by the actions of two peptide hormones, insulin and glucagon.

High blood levels of glucose lead to the increased secretion of insulin, and low blood levels of glucose stimulate the secretion of glucagon. The action of glucagon to stimulate phosphorylase by a cAMP-dependent process has been described (Chapter 7, Section 7.6). The cascade of events involving a series of protein phosphorylations, which is initiated by cAMP as an allosteric effector of protein kinase, leads to a breakdown of glycogen by phosphorylase a. An opposite effect is seen with insulin due to a shift from phosphorylase to synthetase activity, leading to glycogen formation. Glucose has a direct action on phosphorylase a. As a result of binding glucose, a phosphatase is released, which converts phosphorylase a to the inactive b form, and glycogen synthetase D is converted to the

active synthetase I form. The level of glucose thus shifts activity from degradation of glycogen to synthesis. Other binding sites in addition to an allosteric effector site on phosphorylase a for glucose have been considered. It has been proposed that an intracellular mediator of insulin, formed from the insulin–receptor complex, binds to another allosteric site on phosphorylase a to produce an effect similar to that of glucose; that is, removal of phosphate and the formation of the active synthetase I form.

Insulin

Insulin monomer (6K) consists of an A chain of 21 amino acids, and a B chain of 30 amino acids connected by two disulfide bridges (see Figure 19.14 for structure). Aggregates of 2, 4, 6, or more monomers can occur, especially at increased insulin concentrations. The element zinc found with insulin in the pancreas is associated with the aggregate forms of insulin. Insulin is derived from larger precursor peptide molecules, which are synthesized in the endoplasmic reticulum of the β cells of the pancreas. A preproinsulin with a leader signal sequence of 16 amino acids first is cleaved by proteolysis to form the proinsulin molecule, which is extruded into the lumen and stored in the Golgi system. There the proinsulin is gradually cleaved by proteolysis at two points containing arg, arg and arg, lys. The four basic amino acids are released, and a C-peptide (connecting peptide) of 30 amino acids and the intact insulin molecule of 51 amino acids are formed. Insulin and the C-peptide are released from the vesicles due to increased glucose concentrations entering the cell. It is uncertain whether glucose or a glucose metabolite serves as the stimulus, but there is a concomitant increase of Ca^{2+} ion, which triggers insulin release.

Antibodies have been raised against the C-peptide as well as to insulin. Since a C-peptide is released with each insulin molecule, a radioimmunoassay for C-peptide can be used to measure endogenous insulin production. Insulin derived by injection from an exogenous source would not be accompanied by C-peptide, and the low estimates by radioimmunoassay would reflect this difference.

Purified preparations of insulin receptors have been obtained from membrane fractions of liver cells and from fat cells. Solubilized receptor preparations using detergent Triton X-100 are added to an affinity column in which insulin is covalently bound to agarose; the receptor is selectively retained and subsequently eluted with a urea solution. The most purified preparations of receptor indicate a mol wt of 300,000 with subunit size of 125,000, primarily as a dimer (α_2). Since possible small molecular weight proteins may also be present,

the receptor form may be an $\alpha_2\beta_2$ tetramer. It has been suggested that as a result of insulin binding there is first a conformational change, and then a small peptide is released from the receptor, which could be an intracellular mediator of insulin action. Alternatively, insulin has been implicated in the phosphorylation of a 40S ribosomal subunit by an unknown intracellular event that does not involve cAMP activation. It has been estimated that a fat cell contains about 10,000 receptors. The dissociation constant of the insulin–receptor complex is $K_d = 10^{-10}$ M, which corresponds to the level of insulin circulating in blood.

The binding of insulin to its receptor in muscle results in rapid changes in membrane permeability to glucose and amino acids. Two classical experiments are basic to the understanding of insulin action. In the first, intact rat diaphragm muscle is immersed for 10 s in a dilute solution containing insulin, and then washed thoroughly. The treated muscle will incorporate glucose to a far greater extent than its control. In the second experiment, insulin covalently bound to a much larger ferritin molecule, which cannot enter the cell, exerts the same biological effect as soluble insulin. Both experiments demonstrate a rapid effect on membrane function by insulin, and are consistent with the concept that insulin exerts its biological effects while bound to a receptor on the plasma membrane without entering the cell.

Recent experiments suggest that insulin, as well as other peptide hormones such as prolactin and FSH, may be internalized by endocytosis. Radioactively labeled hormone can be localized within the cell lysosomes. This could imply simply a means of inactivation and breakdown by the cell. However, radioactivity has also been localized in the Golgi system, which could be physiologically significant. It is not known if the internalized hormones are biologically active.

Biological Actions of Insulin

Insulin is an important anabolic agent in muscle, liver, and fat cells. Its actions are also anticatabolic. It increases the synthesis of glycogen, fatty acids and triglycerides, and proteins. Insulin achieves these effects by making energy available through the formation of NADPH and ATP. The major source of energy in the cell is produced by glycolysis and subsequent oxidation of the intermediate products via the TCA cycle and oxidative phosphorylation. These events are promoted by the induction by insulin of the key enzymes of glycolysis, phosphofructokinase, and pyruvate decarboxylase. With the lowering of activity levels in liver of fructose 1,6-diphosphatase and pyruvate carboxykinase, it decreases gluconeogenesis.

Glucagon

Glucagon is a straight-chain polypeptide (3.5K) containing 29 amino acids (Figure 16.12). A proglucagon peptide has been isolated. It is secreted by the pancreatic islet α-cell in response to low blood glucose as a stimulus. As with insulin release, a concomitant change in Ca^{2+} ion also is observed. Glucagon is also formed in the cells of the intestine. The principal action of glucagon is to activate the cAMP-dependent phosphorylase system of liver, with a breakdown of glycogen to glucose, which is released into the blood circulation. In addition to its marked effect on glucose production, glucagon increases liver gluconeogenesis by increasing the liver pool of glucose precursors. It decreases pyruvate kinase and acetyl CoA carboxylase activity (Chapter 7, Section 7.5).

In the fat cell glucagon stimulates cAMP-dependent esterase activity to convert triglycerides to free fatty acids and glycerol.

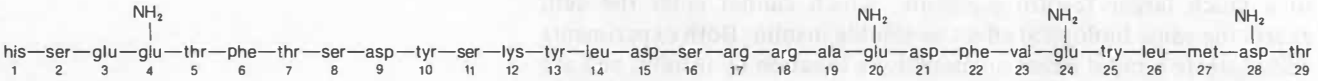


Figure 16.12
Structure of porcine glucagon, a straight-chain peptide containing 29 amino acids.

Somatostatin

In addition to its presence in the hypothalamus, somatostatin has also been isolated in the intestines and the pancreas. It is formed in the islet D cells as a larger peptide precursor and is cleaved to the tetradecapeptide by proteolysis. Somatostatin inhibits the release of glucagon and insulin by the α and β cells, respectively. The inhibition by somatostatin in the pancreas as well as in other tissues, such as the inhibition of release of somatotropin (STH) by the pituitary, appears to be the result of blocking Ca^{2+} entry into the hormone-secreting cells.

Other peptides besides glucagon and somatostatin with known functions found in the gastrointestinal tract include *gastrin* (17 amino acids), which stimulates gastric secretion, *secretin* (27 amino acids), which stimulates the secretion of pancreatic enzymes, and *cholecystokinin* (39 amino acids), which stimulates digestive secretions and gallbladder contractions. It is of interest that the gut contains these as well as a number of other peptides and catecholamines, indoleamines, and endorphins known to be present in the brain.

16.5 HYPOTHALAMIC–PITUITARY HORMONES

For most pituitary hormones, a large peptide precursor is formed by the protein synthesizing system in the endoplasmic reticulum. The protein contains the amino acid sequence for one or several biologically active peptides. The protein precursor may, in fact, contain several species of the same peptide in a single protein sequence. The formation of the biologically active peptide fragments occurs by a process of proteolysis of the parent protein chain, in some cases to form different hormone or prohormone molecules, in other cases to form several copies of the same peptide hormone. Very little is known about the nature of these proteases, what determines the specificity of the cleavage site (amino acid signal sequence), and how the action of the protease is regulated. The larger peptides, identified after chromatographic separation by their immunoreactivity, are referred to as "big," "intermediate," or "little," that is, "big ACTH" or "big" growth hormone, and so on. In the circulating plasma, peptide hormones can aggregate to form dimers or larger molecules and would be referred to as "big" insulin or "big" growth hormone. "Little" insulin would be the monomer-sized molecule (mol wt 6,000).

Endorphins are a class of naturally occurring peptides which bind to morphine receptors (*end*ogenous m*orphin*e-like) in brain tissue. This unexpected development was a direct result of a search for natural substances that could occupy the morphine receptor sites. A physiological relevance for these peptides was indicated on the basis of their ability to mimic the actions of morphine with respect to analgesia, behavior, and withdrawal effects. For this reason the term *opioid* has been used for these peptides. The morphine derivative, naloxone, which competes for the receptor site and blocks morphine effects, also is an antagonist for the endorphins. The first peptides isolated were the enkephalins, containing five amino acids. The most active peptide, containing 30 amino acids, called β-endorphin, was first isolated from β-lipotropin (see Figure 16.16).

There are no direct nerve connections from the hypothalamus to the anterior pituitary, but there is a rich blood supply from the hypophyseal portal system. When the pituitary stalk is transected, serum prolactin levels rise, while the levels of other anterior pituitary hormones fall. This demonstrates the stimulatory effect of the hypothalamus on the anterior pituitary hormone secretions and the dominant inhibitory effect on the control of prolactin secretion. (See Figure 16.13.) Hypothalamic actions are mediated by chemical

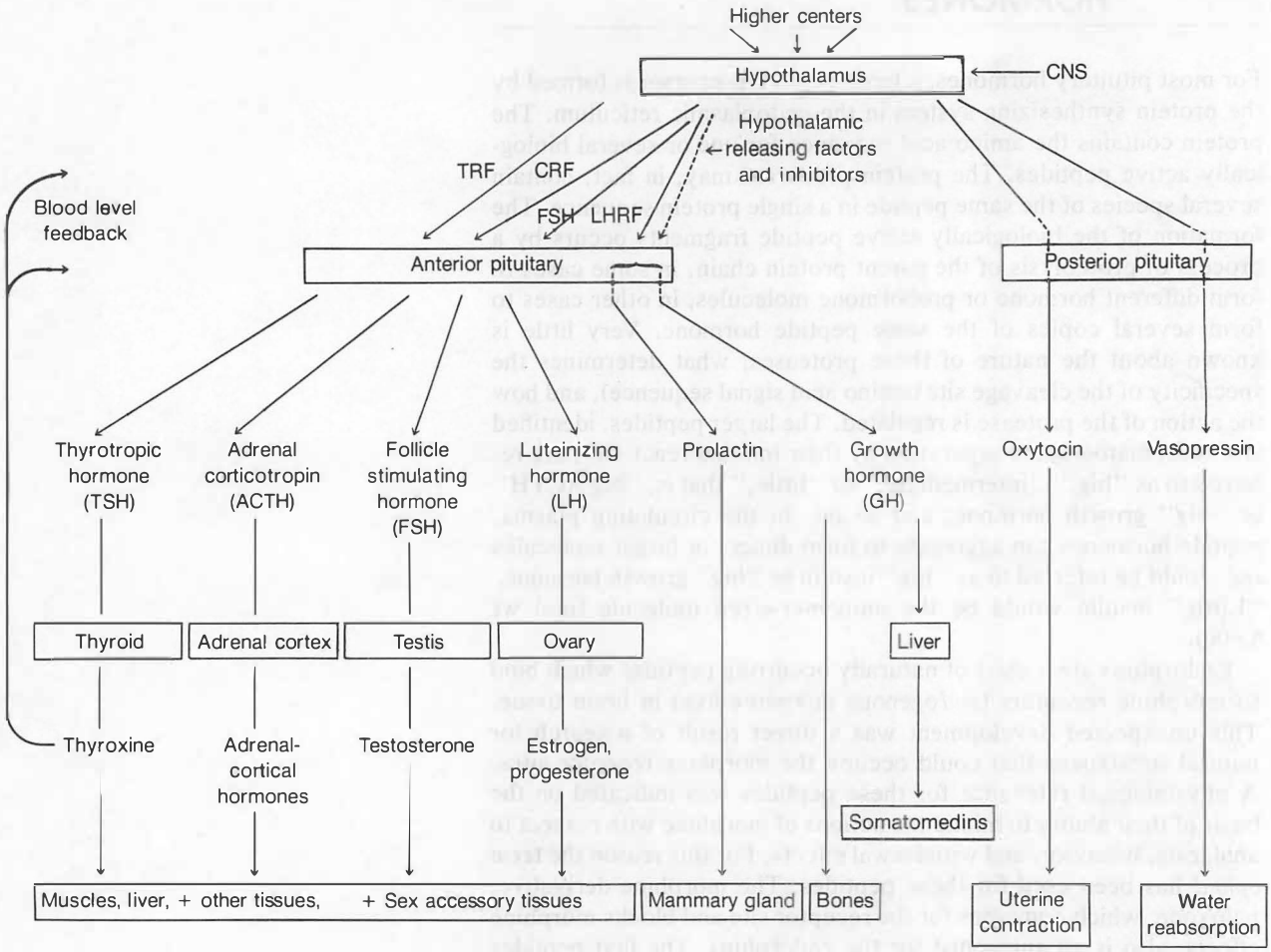


Figure 16.13
Hypothalamus-pituitary-target organ relationships.
Releasing factors from the hypothalamus stimulate pituitary trophic hormones, which stimulate target organs to secrete their respective hormones. These target organ hormones secreted into the circulation will exert a negative feedback at the level of the hypothalamus and pituitary tissues to inhibit further secretion of releasing factors and trophic hormones. Prolactin and growth hormones have a release-inhibiting factor as well as a releasing factor. Oxytocin and vasopressin are produced in the hypothalamus and are stored in the posterior pituitary.

messengers, which are transported by the portal blood system from the median eminence of the hypothalamus to the anterior pituitary. By contrast, the posterior pituitary has direct neuronal connections to hypothalamic nuclei. The hormones vasopressin and oxytocin, which are secreted from the posterior pituitary gland, are synthesized in cells of the hypothalamic nuclei, travel down the nerve axons, and are stored in vesicles in the posterior pituitary. Other regulatory agents or hormones are produced in the hypothalamus by these peptidergic neurons, with cell bodies located in different hypothalamic nuclei. The activity of these neurons are controlled by other agents such as the neurotransmitters, dopamine, norepinephrine, and serotonin.

Hypothalamic Releasing Factors

Three of the hypothalamic peptides have been characterized and synthesized. They are (1) thyrotropin releasing factor (TRF), (2) the gonadotropin releasing factor (GnRF), and (3) the growth hormone release inhibiting factor (GRIF), also called somatostatin (Figure 16.14). Biological activity in portal blood indicates the presence also of separate releasing factors for growth hormone (GHRF), prolactin (PRF), and corticotropin (CRF); however, their chemical nature remains undetermined. The release of pituitary trophic hormones is inhibited by a negative feedback control by the target organ hormones acting on hypothalamic and pituitary secretory cells. A retrograde blood flow in the pituitary stalk can deliver the pituitary hormones to the median eminence and result in a "short loop" regulation of the secretion of hypothalamic releasing factors. In particular, by exerting both negative and positive feedback at the hypothalamic level, prolactin and growth hormone may regulate their own rate of production by altering the secretion of releasing factors or release-inhibiting factor.

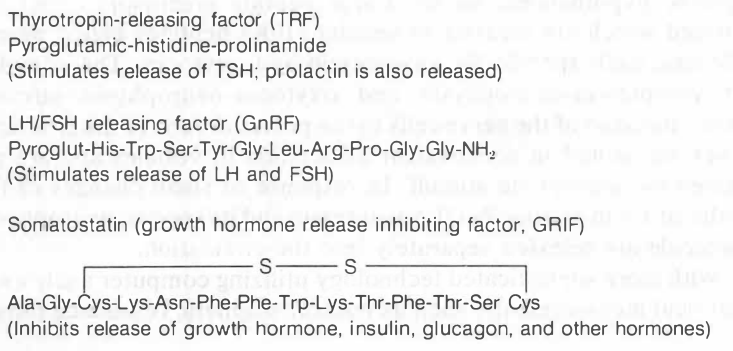


Figure 16.14
Hypothalamic releasing factors.

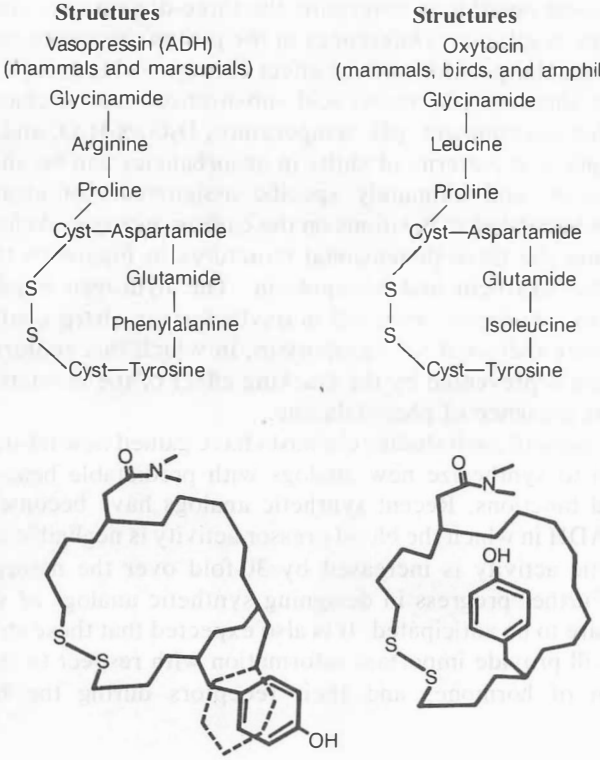
Oxytocin and Vasopressin

Two of the peptides present in the posterior lobe of the pituitary, oxytocin and vasopressin, are shown in Figure 16.15, which also lists their major biological actions. Oxytocin, named for its action on uterine tissue to cause contraction of the myometrium, also has an effect on the smooth muscle of the ducts of the mammary gland to cause the expulsion of milk. Oxytocin is released in response to CNS signals in response to the suckling stimulus. Vasopressin, named for its action to increase blood pressure, is also known as the antidiuretic hormone (ADH) for its action at the distal convoluted tubule of the kidney, which results in reabsorption of water.

These two hormones are important due to their prominent role in the history of the development of our biochemical concepts of biologically active peptides. They were the first to be isolated and purified, the first to be synthesized. A number of analogs with varying degrees of biological activity, which could be attributed to specific changes in amino acid structure, were synthesized during this period. A Nobel Prize was awarded in 1954 to Vincent Du Vigneaud for these studies. It was demonstrated that a single peptide structure could have more than one biological action. The effect on biological activity varied according to the type of amino acid substitution and its position in the molecule. Whereas the substitution of leucine (nonpolar) for arginine (basic) results in a profound change in activity (Figure 16.15), the substitution of one basic amino acid (arginine, in human ADH) for another (lysine, in ovine ADH) does not alter the type or degree of biological action. In some cases, the substitution of leucine for arginine, for example, confers new properties not observed in either vasopressin or oxytocin. The substance called vasotocin was first synthesized before it was subsequently found to be a natural peptide present in most nonmammalian species. The new property, maintaining water balance, is particularly important in amphibia.

Oxytocin and vasopressin are synthesized in neuronal cells of specific hypothalamic nuclei. Large peptide precursors (20K) are formed which are cleaved to smaller (10K) peptides called neurophysins, each specific for vasopressin and oxytocin. The complex of vasopressin–neurophysin and oxytocin–neurophysin streams down the axon of the nerve cells to the posterior lobe of the pituitary. They are stored in noncovalent association in vesicles and are released by appropriate stimuli. In response to small changes of the order of 1% in plasma [Na^+], vasopressin and its specific neurophysin molecule are released separately into the circulation.

With more sophisticated technology utilizing computer analysis of physical measurements, such as nuclear magnetic resonance (nmr),



Biological Activity

	Vasopressin	*Oxytocin*
Antidiuretic activity	++++	+
Blood pressor activity (rat)	++++	+
Uterine contraction (rat)	+	++++
Milk ejection	+	++++

++++ = very active in 10^{-9}–10^{-14} g range; + = <5% activity.

Figure 16.15
Structures and biological actions of vasopressin and oxytocin.
The middle figures depict three-dimensional (solution) structures obtained from nmr data. Phenylalanine stacking with tyrosine prevents H-bonding of tyrosine to amide of asparagine that occurs in oxytocin.
From R. Walters, *Fed. Proc.*, 36:1872, 1977. Reproduced with permission.

it has become possible to determine the three-dimensional structure of peptides in solution. Differences in the proton resonance patterns due to molecular position and the effect of neighboring groups can be related to alterations in amino acid substitutions and in changes in the solution environment, pH, temperature, D_2O vs H_2O, and so on. The complicated patterns of shifts in absorbancies can be analyzed by computer, and ultimately specific assignments of absorption peaks can be related to positions on the carbon–nitrogen skeleton. In this manner the three-dimensional structures in Figure 16.15 were derived for oxytocin and vasopressin. The hydrogen bonding of tyrosine to asparagine observed in oxytocin is in sharp contrast to the structure indicated for vasopressin, in which this conformation by tyrosine is prevented by the stacking effect of the aromatic rings due to the presence of phenylalanine.

On the basis of such studies chemists have gained new information by which to synthesize new analogs with predictable behavior in biological functions. Recent synthetic analogs have become available for ADH in which the blood pressor activity is negligible and the antidiuretic activity is increased by 30-fold over the natural substance. Further progress in designing synthetic analogs of specific function are to be anticipated. It is also expected that these structure studies will provide important information with respect to the conformation of hormones and their receptors during the binding process.

Relationship of ACTH to Opioid Peptides

The role of ACTH in adrenal steroidogenesis is described in Chapter 15, Section 15.3. The biosynthesis of ACTH and a family of related peptides, including α-melanocyte stimulating hormone (α-MSH) and the endorphins, are discussed in this section.

ACTH is a straight-chain peptide (mol wt 4,500) containing 39 amino acids, without disulfide bridges (Figure 16.16). The mammalian ACTH structures that have been sequenced vary by only two amino acids in the carboxyl terminal half of the molecule. The amino terminal half contains the steroidogenic activity. The molecule has been synthesized, and peptides containing the first 18 to 24 amino acids have activity. Commercial preparations containing 24 amino acids are used clinically and are equal in most respects to the natural ACTH of 39 amino acids. The carboxyl terminal half (ACTH, 18–39) is referred to as CLIP and is devoid of biological activity. Radioimmunoassays are available using antibodies raised against the ACTH (1–18), ACTH (18–39), as well as the complete molecule of ACTH (1–39). The melanocyte stimulating hormone, α-MSH, is the amino terminal half of ACTH (1–13) with Ser^1 in the acetylated form. With

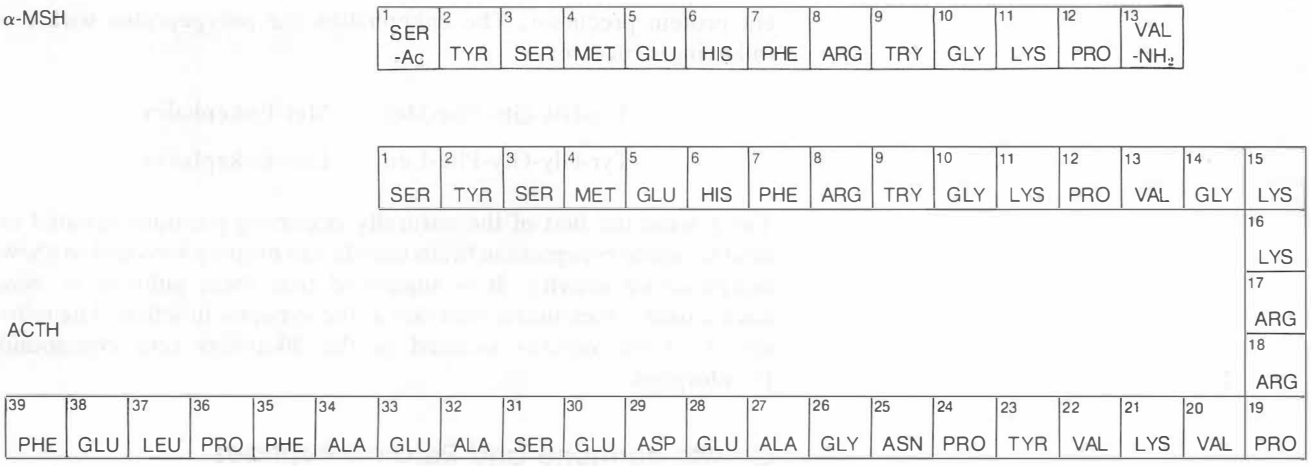


Figure 16.16
Amino acid homology of α-MSH and ACTH.
α-MSH has an acetyl group on Ser[1]. It has no ACTH activity. ACTH has α-MSH-like effect on increasing skin pigmentation. ACTH action requires the presence of amino acids 1–18. The fragment containing amino acids 18–39 is known as CLIP. It has no biological activity.

the use of immunoassay and immunoprecipitation techniques, it has been possible to relate those forms to larger peptide molecules present in the anterior pituitary and median eminence, which could serve as biosynthetic precursors. With these techniques it has been possible to show that large peptides of 31K ("big" ACTH), 13K ("intermediate" ACTH), and 4.5K ("little" ACTH) contain amino acid sequences in common with the ACTH (1–39) molecule. Similar studies using antibodies to β-lipotropin and β-endorphin indicate a common recognition site in the 31K peptide. The 31K peptide, which reacts with both the antibodies to ACTH (1–39) and β-lipotropin contains these two separate amino acid sequences and is referred to as pro-ACTH/endorphin. The complete analysis of this precursor peptide and the relationship of the identifiable fragments to the entire molecule is shown in Figure 16.17.

The 31K peptide contains a 16K fragment not yet identified with any known biological activity, ACTH (1–39), the β-LPH (1–91), the carboxyl end of which contains β-endorphin (61–91). Although the latter fragment also contains the five amino acid sequence of methionine–enkephalin, it is likely that this pentapeptide, which has been found in a variety of tissues, is formed from an entirely differ-

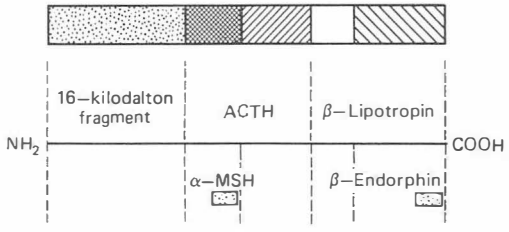


Figure 16.17
The 31K peptide–pro-ACTH/endorphin peptides.
The 31K peptide contains the fragments α-MSH (ACTH 1–13), the ACTH (1–39) molecule, β-endorphin (β-lipotropin 61–91) and β-lipotropin (1–91) molecule. The first 16K fragment is devoid of activity.

ent protein precursor. The enkephalins are polypeptides with the following structures:

Tyr-Gly-Gly-Phe-Met	Met-Enkephalin
Tyr-Gly-Gly-Phe-Leu	Leu-Enkephalin

These were the first of the naturally occurring peptides isolated to bind to opiate receptors in brain membrane preparations and to show morphinelike activity. It is suggested that these substances may have a neural transmitter function at the synaptic junction. The most active of the opiates isolated is the 30-amino acid compound β-endorphin.

Growth Hormone and Related Peptides

Growth hormone and prolactin are straight-chain polypeptides that are similar in size and structure and share in some of their biological properties (Table 16.1). In addition there are proteins produced by the placenta that have growth hormone and prolactinlike properties. These substances, called placental lactogens, can be differentiated from the pituitary hormones by the use of hormone-specific antibodies. Due to similarity in structures, antibodies generated for each specific hormone will cross-react to some extent with the other two protein hormones. (See Figure 16.18.)

A variety of factors have been implicated in growth hormone regulation. Growth hormone release is regulated by a releasing factor and by SRIF (somatostatin). The highest concentration of somatostatin is found in the hypothalamus, but it has been found in other parts of the central nervous system and in other tissues such as the pancreas. The structure of the specific releasing factor is unknown. The neurotransmitters dopamine and norepinephrine have known

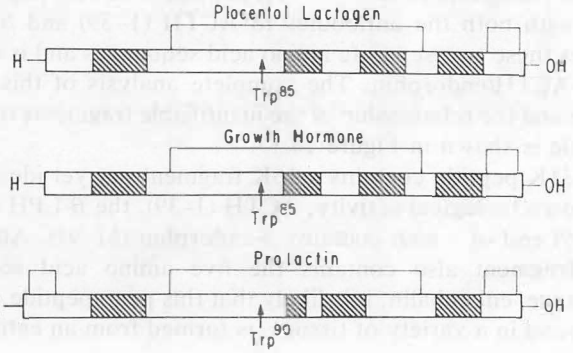


Figure 16.18
Similarity in structures of prolactin, growth hormone, and placental lactogen.
Homology indicated by position of disulfide bridges (thin lines), repeating polypeptide sequences internally, and position of tryptophan.
From H. Niall, M. Hogan, R. Sauer, I. Rosenblum, and F. Greenwood, *Proc. Natl. Acad. Sci., USA.* 68:869, 1971. Reproduced with permission.

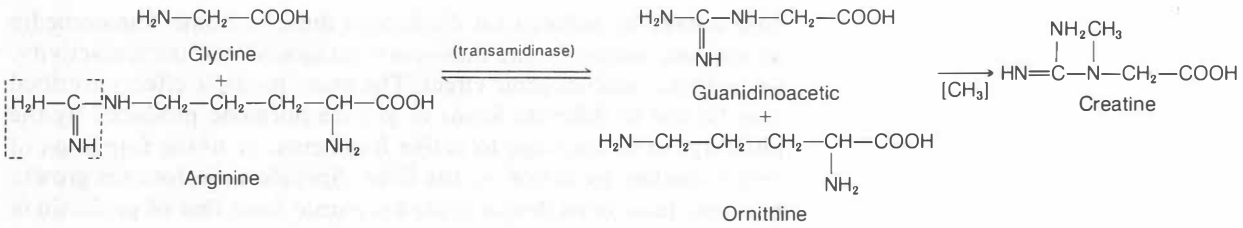


Figure 16.19
Synthesis of creatine.

stimulatory effects. Hypoglycemia, exercise, fasting, and amino acids, particularly arginine, are some factors that stimulate the release of growth hormone. As its name implies, growth hormone is active in the regulation of a number of growth processes. The term somatotropin (STH), by which GH is also known, connotes a wider range of activity, which affects soft tissues, organs, and bones. It is well to remember that there are other agents such as insulin and thyroid hormone that also exert important stimulatory effects during the period of growth. In particular, to achieve normal growth rates thyroid hormone and growth hormone act synergistically during development. Thyroid hormone may be necessary for the optimum production of growth hormone in the pituitary. Both are involved in achieving normal rates of protein synthesis in the liver during growth, and in the induction of the enzyme transamidinase in the kidney. In the hypophysectomized animal, growth hormone is specifically required for the formation of guanidinoacetate, which is the direct precursor of creatine (Figure 16.19). These reactions occur in the kidney and liver, respectively.

Growth hormone can be considered a prohormone, since many of its actions are mediated through peptide intermediates called somatomedins. Three somatomedins, A, B, and C, have been characterized. The term somatomedin refers to plasma factors, which mediate growth and are dependent on growth hormone. It is not known whether they are derived from larger growth hormone peptides by proteolysis or if they are synthesized under stimulus of growth hormone. This should be resolved when their amino acid sequences are determined and compared to that of growth hormone. Other factors that affect growth of tissues, such as nerve growth factor and epidermal growth factor, are formed in liver but do not require stimulation by growth hormone.

One of the somatomedins, formerly referred to as *sulfation factor,* stimulates sulfate uptake by cartilage. The incorporation of $^{35}SO_4$ is used as a bioassay for growth hormone or somatomedin. Somatomedins also stimulate protein synthesis and amino acid and glu-

cose uptake by isolated rat diaphragm muscle. Some somatomedin actions are similar, while others are antagonistic to insulin activity, for example, diabetogenic effect. The many multiple effects ascribed may be due to different forms of growth hormone produced by the pituitary, to its cleavage to active fragments, or to the formation of somatomedins by action on the liver. Specific receptors for growth hormone have been demonstrated separate from that of prolactin in liver, kidney, and in other tissues.

Differences in peptide structure among animal species are greater for growth hormone than for the other anterior pituitary hormones. Only primate growth hormone, monkey or human, is active in the human. Bovine growth hormone is active in murine and avian species, and fragments after proteolytic procedures have some activity in humans. Different portions of the molecule that have been synthesized contain some biological activity. For these reasons it is felt that some smaller peptide fragment (an active core) may be involved in expression of biological activity.

Human growth hormone has 191 amino acids (Figure 16.20) with mol wt 21,500 and pI 4.9. There are two disulfide bridges in a straight chain on purification. It can form aggregates and has been assayed and isolated from plasma as a dimer. Due to the heterogeneity of circulating growth hormone and the presence of somatomedins, the extent of immunoreactivity of plasma constituents by RIA for growth hormone is quite variable and often much lower than estimates based on radioreceptor binding assays or assays for biological activity.

Prolactin

Prolactin is a hormone found throughout the animal kingdom and has numerous and diverse effects that regulate a variety of physiological functions. These activities are grouped into actions related to water and electrolyte metabolism, actions related to reproduction, actions affecting growth, developmental, and metabolic events. In mammals prolactin has osmoregulatory functions and growth-promoting effects, but the best-known activity in the human is related to the growth of the mammary glands and the initiation of lactation. In the rodent species, prolactin stimulates corpus luteum function and modulates growth and secretions of male sex accessory organs. Placental lactogen is a substance produced during pregnancy, which shares in having attributes of both pituitary prolactin and pituitary growth hormone. An older name for this substance is chorionic somatomammotropin. As with growth hormone, the actions of prolactin are complicated by the fact that they are synergistic to or are modulators of other hormonal agents. Prolactin affects the release of

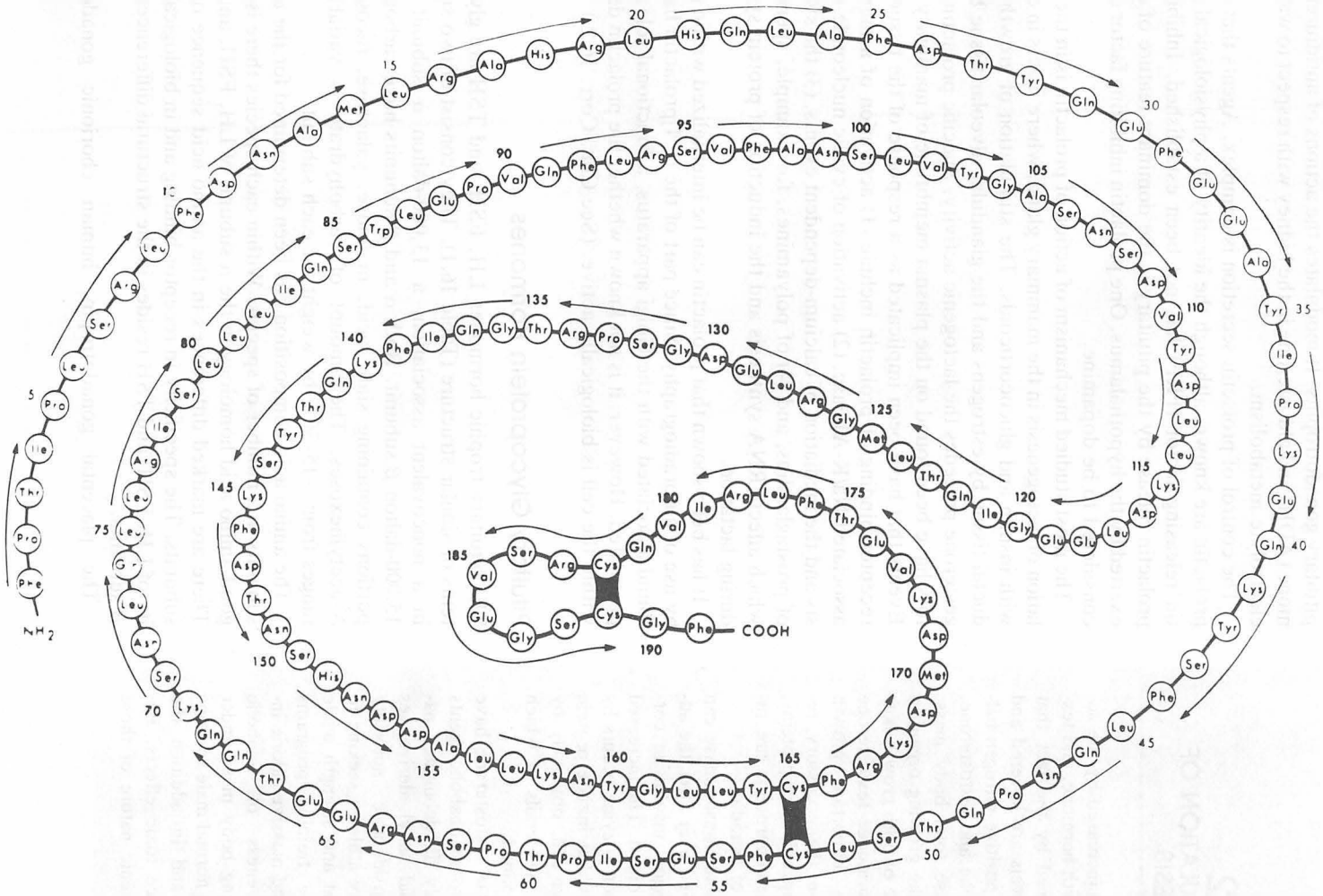


Figure 16.20
Structure of human growth hormone.
The hormone contains 191 amino acids with two disulfide bridges.
From C. H. Li and T. H. Bewley, *Proc. Natl. Acad. Sci. USA*. 73:1476, 1976.
Reproduced with permission.

CLIN. CORR. 16.2 HORMONE REGULATION OF PROTEIN SYNTHESIS

Protein anabolism is increased by the actions of insulin, growth hormone, and testosterone as evidenced by the fact that blood amino acid levels are lowered and that they produce a positive nitrogen balance. Growth hormone and testosterone, in particular, increase total body mass. Optimal growth in the growing organism requires the presence of both growth hormone and thyroid hormone. Testosterone and estradiol have important growth promoting effects on the accessory reproductive tissues. Specific liver proteins, CBG, TBG, and metalloproteins are induced by the action of estradiol.

The enzyme tyrosine transaminase can be induced in perfused liver after the administration of glucagon, insulin, or cortisol or by a protein digest. The increased synthesis of a common enzyme occurs by the actions of agents that initiate protein synthesis in the liver cell, obviously by separate pathways, the details of which have yet to be determined.

Synthetic analogs of testosterone have been developed as protein anabolic agents for use in chronically ill, debilitated patients. These steroid alkyl derivatives have received considerable notoriety owing to their use by athletes seeking to increase body weight and strength while undergoing intensive training programs involving exercise and massive caloric intake. The effectiveness of anabolic steroids for increasing body mass under these conditions in a normal male remains to be established, and in addition the steroids may produce toxic effects. Because of the androgenic nature of these

pituitary gonadotropins. It modulates the actions of antidiuretic hormone (ADH) and aldosterone in the kidney with respect to water and electrolyte metabolism.

The control of prolactin secretion is complex. Agents that release prolactin are known, although the identity of a physiological prolactin releasing factor (PRF) has not been established. Inhibition of prolactin release by the pituitary is the dominant feature of control exerted by the hypothalamus. One prolactin inhibiting factor (PIF) is considered to be dopamine.

The most studied mechanism of action of prolactin is in the stimulation of lactogenesis in the mammary gland, where it acts in concert with insulin and glucocorticoids. The stimulation of growth of the ductal tissue by estrogens and the glandular alveolar tissue by progesterone precedes the lactogenic activity. Specific prolactin receptors have been found on the plasma membrane of mammary tissue. Events that have been implicated as a response of the tissue to the receptor binding of prolactin include (1) activation of a membrane associated Na/K-ATPase, (2) activation of cyclic nucleotide synthesis and the mediation of calcium-dependent events, (3) the synthesis of prostaglandins, and (4) of polyamines, for example, spermidine, which affects RNA synthesis and the induction of protein synthesis during lactation.

It has been shown that prolactin can be internalized within the cell by use of autoradiography. Since part of the [^{125}I]prolactin has been found associated with the Golgi apparatus, a functional role can be postulated. However, it is not known whether the prolactin detected within the cell is biologically active. (See Clin. Corr. 16.2.)

Pituitary Glycoprotein Hormones

The pituitary trophic hormones LH, FSH, and TSH are glycoproteins of similar structure (Table 16.1). They consist of two subunits in a noncovalent association, a 13,000-dalton α subunit and a 15,000-dalton β subunit. Both α and β subunits have carbohydrate portions containing sialic acid, mannose, galactose, fucose, and *N*-acetylhexoses. The amount of carbohydrate is variable and ranges from ~15–30% by weight for each subunit.

The amino acid composition has been determined for the α and β subunits in a number of species. Within each species there is a very good amino acid homology in the α subunit of LH, FSH, and TSH. There are marked differences in the amino acid sequence of the β subunits. The specificity in receptor binding and in biological activity of LH, FSH, and TSH resides in the structural differences of the β subunits.

The placental gonadotropin, human chorionic gonadotropin

(HCG), although not of pituitary origin, is structurally and biologically very similar to the luteinizing hormone (LH). Because of its availability from pregnancy urine, HCG has been used clinically instead of the more difficult to obtain LH. HCG also contains an α subunit very similar to that of the pituitary hormones, but the β subunit is markedly different from that of FSH and TSH. The amino acid sequences of the β subunit of HCG and LH are sufficiently alike to account for the similar biological activities of the two hormones. HCG and LH can be distinguished immunologically, however.

Antibodies can be produced that are relatively specific for LH, FSH, TSH, and HCG by injection of pure trophic hormone of one species into another. The specificity of the antibody resides primarily in the differences in amino acid structure of the β subunit. Due to the similarity in amino acid sequences of the α subunit, antibodies to the entire protein molecule interact to some extent (cross-react) with the other trophic hormones. Both α and β subunits of the pituitary hormones can be obtained in pure form; indeed, they are present in the tissues in subunit form to a limited extent where they have been isolated. The separate subunits have little biological activity, but

agents, their use by female athletes produces more obvious masculinizing effects. It is now routine practice to test for steroids as well as other drugs in the urine prior to major athletic events.

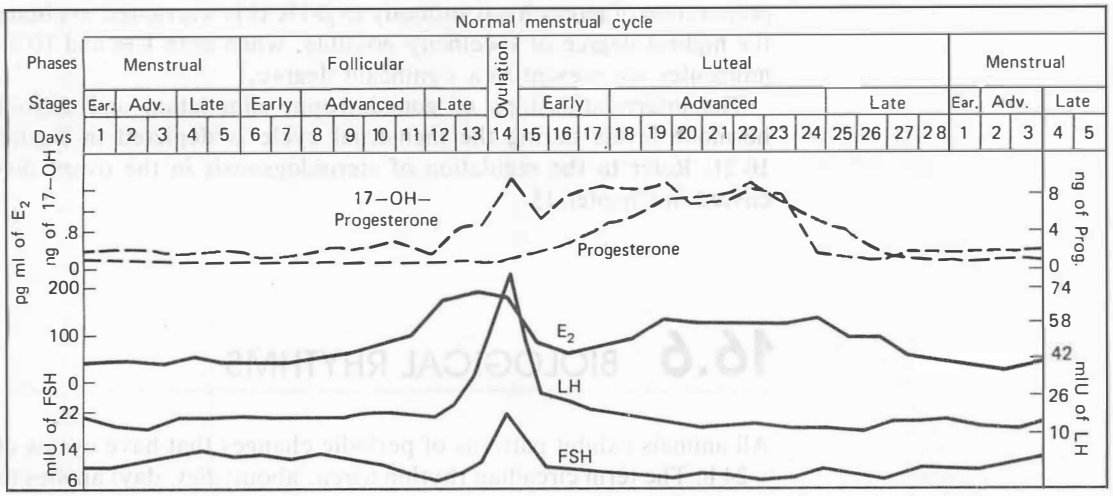


Figure 16.21
Human menstrual cycle.
The relationships among the gonadotropins, LH and FSH, and the steroid hormones, estradiol and progesterone, at the time of ovulation should be noted. Refer to Chapter 15 on ovarian steroidogenesis.
From G. T. Ross, C. M. Cargille, M. B. Lipsett et al., *Recent Prog. Hormone Res.*, 26:1, 1970.

they are immunologically active. Antibodies raised against the respective pure α and β subunits have the expected specificities; the α-antibodies will react with each hormone, LH, FSH, TSH, and HCG, whereas the β-antibody is quite specific for only the one entire hormone (or its β subunit) from which it was derived. For example, there is no cross-reactivity between the antibody of the β subunit of HCG for FSH or TSH. Since there are similarities between the β subunits of LH and HCG, some cross-reactivity can be observed. Generally, the small degree of cross-reactivity of antibodies from different hormones or subunits may not present a problem for routine purposes of measurement, that is, assay for circulating concentrations of pituitary hormones in blood.

An antibody to any one α-peptide will provide the needed information as well as another. However, in certain conditions the small degree of cross-reactivity may become significant: as in the measurement of HCG in early pregnancy, the monitoring of ectopic hormone-producing tumors, or in the preparation of a vaccine against pregnancy. In each case, cross-reaction with pituitary LH would be misleading. The use of the purest form of β-HCG subunit or the use of a chemically modified β subunit as an antigen, or the preparation of monoclonal antibody to β-HCG is warranted to obtain the highest degree of specificity possible, when both LH and HCG molecules are present to a significant degree.

The interrelationships of gonadotropin stimulation and steroid hormone levels during the menstrual cycle is depicted in Figure 16.21. Refer to the regulation of steroidogenesis in the ovary discussed in Chapter 15.

16.6 BIOLOGICAL RHYTHMS

All animals exhibit patterns of periodic changes that have cycles of ~24 h. The term circadian rhythm (*circa*, about; *dies*, day) applies to rhythms that persist in the absence of any environmental or sensory input that are not exactly 24 h in length; they are free-running and can be compared to autonomous oscillations. In fact, however, most rhythms are entrained by external periodic signals such as light and dark, temperature changes, feeding, and activity patterns. These internal and environmental factors are referred to as zeitgebers (time keepers). If a human is shielded completely from all possible input (men have been isolated alone in deep mines underground with no outside stimuli or time cues for several months) their circadian

rhythms free-run at their own frequency of slightly less than or longer than 24 h. In the human there is a multiplicity of oscillators, which may be coupled or desynchronized and free-run with different frequencies. The search for the biochemical or physiological equivalent of an internal oscillator, in spite of continued research, has not been successful.

Desychronization of rhythms can have profound physiological effects. Best known of these is the condition of "jet-lag," which occurs when many time zones are crossed in a jet plane, particularly in traveling west to east. Recovery requires several days, and usually a week is needed to reestablish normal circadian rhythms.

The endocrine system provides a good model for the study of circadian rhythms, since they have pronounced cycles with fluctuations of hormone levels that are often greater than 50% during the light/dark cycle. A well-established cycle is that of corticosteroid levels in blood or urine, which is a consequence of the stimulation by hypothalamic (CRF) and pituitary (ACTH) hormones (Figure 16.22). The degree of inhibitory negative feedback exerted by cortisol will vary at different phases of the 24-h cycle. These effects have practical clinical implications with respect to the timing of the administered dose of cortisol. Furthermore, the efficacy in response to drugs in general can be altered, depending upon the time at which they are administered. The normal cortisol cycle in the human peaks in the early AM and has its lowest values in late afternoon. As a result of hypersecretion in Cushing's disease, cortisol values are high in both PM and AM. The lack of a cycle, due to relatively high cortisol values in the PM as well as in the AM, has been used as a diagnostic test for adrenal hyperactivity of Cushing's.

Another organ that elicits a pronounced circadian response is the pineal gland. A number of biologically active derivatives of tryptophan, including serotonin and melatonin, are produced in the pineal. In some species, as in the amphibia, external light serves as a direct stimulus to pineal activity. In other species, including the human, the pineal stimulation is through release of neurotransmitters by nerve impulses that emanate from other nerve centers, including the optic nerve. Melatonin and other pineal secretions have effects on the hypothalamus–pituitary system. There is a pronounced circadian rhythm in the formation of melatonin, since its formation occurs only in the dark. The sharp peak in activity in the dark is due to the activity of a single enzyme, *N*-acetyltransferase, which converts serotonin to acetylserotonin (Figure 16.23). Neither serotonin formation nor the *O*-methyltransferase, which converts acetylserotonin to melatonin, is affected by light/dark cycles. In many rhythms the peak and trough of activity correspond to the temporal behavior of the animal; the nocturnal animal has a peak in corticosteroids in the

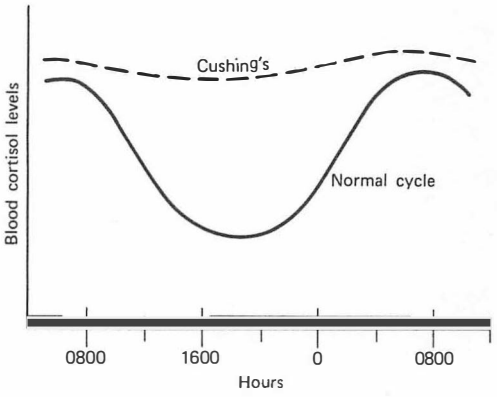


Figure 16.22
Absence of normal AM/PM variation in cortisol levels in Cushing's disease.

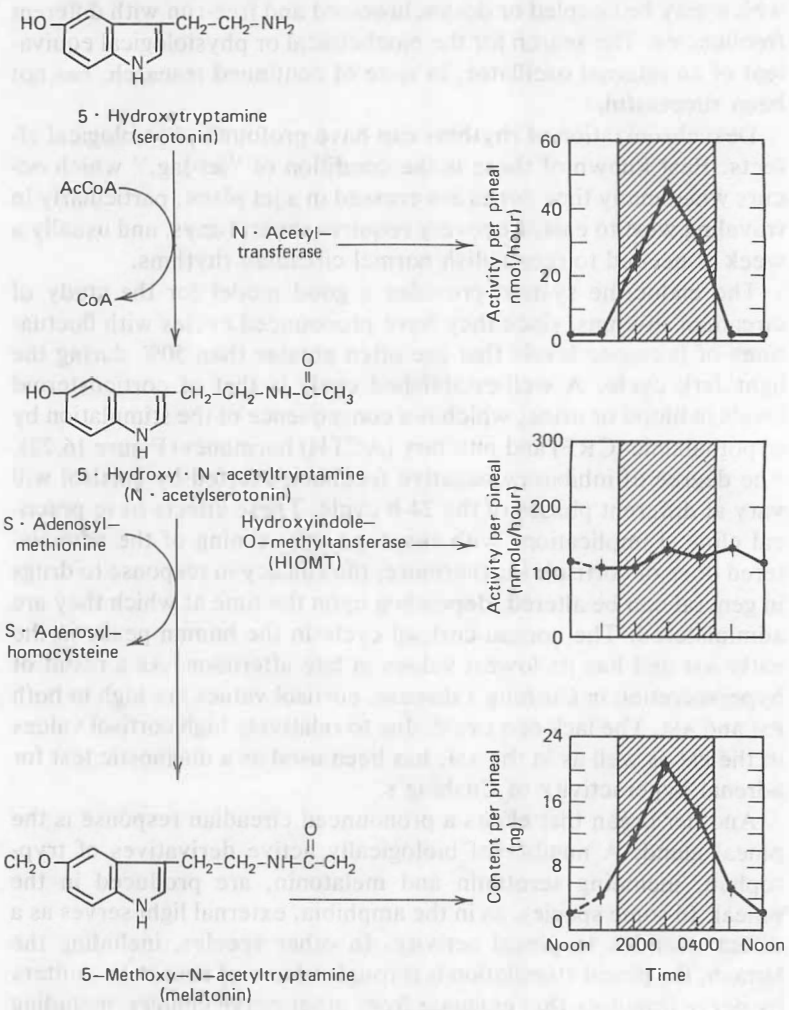


Figure 16.23

Circadian rhythm in the formation of melatonin in the pineal gland of the chick is due to the stimulation of N-acetyltransferase in the dark.

The O-methyltransferase (HIOMT) as well as the formation of precursor serotonin do not have large activity changes in the light/dark cycle.

From S. Brinkley, S. MacBride, D. Klein, and C. Ralph, *Science*, 181:273, 1974. Reproduced with permission. Copyright 1974 by the American Association for the Advancement of Science.

PM, and the diurnal animal (human) has its corticosteroid peak in the early AM. In contrast, the melatonin peak formation occurs in the dark in early AM regardless of activity periods. Pineal and melatonin activity have been implicated in the regulation of gonad and adrenal function. The pineal is considered a part of the chain that determines phase relationships, with respect to circadian rhythms of these endocrine units. It is not likely that the pineal is the basic pacemaker of the body; it could be responding to signals from higher centers, for example. It is probable that there are a number of pacemakers, since individual cells may have endogenous clocks of their own; although a pacemaker could be more dominant in one species than another in determining certain patterns of activity or behavior.

BIBLIOGRAPHY

Cell Regulation

Goldstein, J. L., Anderson, R. G. W., and Brown, M. S. Coated pits, coated vesicles, and receptor-mediated endocytosis. *Nature* 279:679, 1979.

Greengard, P. Phosphorylated proteins as physiological effectors. *Science* 199:146, 1978.

Means, A. R. and Dedham, J. R. Calmodulin: An intracellular calcium receptor. *Nature* 285:73, 1980.

Schlessinger, J. The mechanism and role of hormone induced clustering of membrane receptors. *Trends Biochem. Sci.*, 5:210, 1980.

Prostaglandins

Gorman, R. J. Modulation of human platelet function by prostcyclin and thromboxane A_2. *Fed. Proc.* 38:83, 1979.

Insulin

Drazmin, B., Solomons, C. C., Toothaker, D. R., and Sussman, K. E. Energy dependent steps in insulin-hepatocyte interactions. *Endocrinology*, 108:8, 1981.

Kasvinsky, P. J., Shechosky, S., and Fletterick, R. J. Synergistic regulation of phosphorylase a by glucose and caffeine. *J. Biol. Chem.* 253:9102, 1978.

Oxytocin

Walter, R. Identification of sites in oxytocin involved in uterine receptor recognition and activation. *Fed. Proc.* 36:1872, 1977.

Hormone Endocytosis

Josefsberg, Z., Posner, B. I., Patel, B., and Bergeron, J. J. M. The uptake of prolactin into female rat liver. *J. Biol. Chem.* 254:209, 1979.

Pilch, P. F., and Czech, M. P. The subunit structure of the high affinity insulin receptor. *J. Biol. Chem.* 255:1722, 1980.

Prolactin

Rillema, J. A. Mechanism of prolactin action. *Fed. Proc.* 39:2593, 1980.

Growth Hormone

Martin, J. B., Functions of central nervous system neurotransmitters in regulation of growth hormone. *Fed. Proc.* 39:2902, 1980.

Paladini, A. C., Pena, C., and Retegni, L. A. The intriguing nature of the multiple actions of growth hormone. *Trends Biol. Sci.* 4:250, 1979.

Gonadotropins

Ghai, R. D., Mise, T., Pandian, M. R., and Bahl, D. P. Immunological properties of the β-subunit of human chorionic gonadotropin. *Endocrinology* 107:1556, 1980.

ACTH/Endorphins

Eipper, B. A., and Mains, R. C. Structure and biosynthesis of proadrenocorticotropin/endorphin and related peptides. *Endocrine Rev.* 1:1, 1980.

Biological Rhythms

Deguchi, T. Ontogenesis of a biological clock for serotonin:acetyl coenzyme A *N*-transferase in the pineal gland of the rat. *Proc. Natl. Acad. Sci. USA* 72:2814, 1975.

Spelsberg, T. C., and Halberg, F. Circannual rhythms in steroid receptor concentration and nuclear binding in chick oviduct. *Endocrinology* 107:1234, 1980.

17

DNA

STELIOS AKTIPIS

THE REPLICATIVE PROCESS AND REPAIR

17.1 BIOLOGICAL PROPERTIES OF DNA

Introduction

This chapter reviews the chemical structure of DNA and examines the relationship between this structure and the biochemical function of DNA. Within this context the process of DNA replication and repair is detailed. The remaining processes through which DNA regulates the expression of biological information (i.e., transcription and translation) are the subjects of other chapters.

One of the striking aspects of natural order is the sense of unity that exists between the members of successive generations in each species. It is apparent that an almost totally stable bank of information must always be preserved and passed from one generation to the next if individual species are to maintain their identities relatively unchanged over millions of years. It is now well established that the bank of genetic information takes the form of a stable macromolecule, deoxyribonucleic acid (DNA), which serves as the carrier of genetic information in both procaryotes and eucaryotes. DNA exhibits a rare purity of function by being one of the few macromolecules known to perform, with only minor exceptions, the same basic functions across species barriers.

It is apparent that the properties of cells are to a large extent determined by their constituent proteins. Many proteins serve as indispensable structural components of the cell. Other proteins, such as enzymes and certain hormones, are functional in character and determine most of the biochemical properties of the cell. As a result, the factors that control *which proteins* a cell may synthesize, at *what quantities,* and at *which sequence* are the same factors that primarily determine the function as well as the destiny of every living cell.

It is now well recognized that DNA is the macromolecule that ultimately controls, primarily through protein synthesis, every aspect of cellular function. DNA exercises this control as suggested by the sequence

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{protein}$$

The flow of biological information is clearly from one class of nucleic acid to another, from DNA to RNA, with only minor exceptions, and from there to protein. In order for this transfer of information to occur faithfully, each preceding macromolecule serves as a structure-specifying template for the synthesis of the subsequent member in the sequence.

In addition to regulating cellular expression, DNA plays an exclusive role in heredity. This role is suggested by a circular arrow engulfing DNA, which depicts DNA as a *replicon*, a molecule that can undergo self-replication. The significance of *replication* is far-reaching. It permits DNA to make copies of itself as a cell divides. These copies are bestowed to the daughter cells, which can thus inherit each and every property and characteristic of the original cell.

First, the important message to be retained is that DNA ultimately determines the properties of a living cell by *regulating the expression of biological information*, primarily by the control of protein synthesis. Second, but not less importantly, it should be clear that DNA transfers biological information from one generation to the next, that is, it is essential for the *transmittance* of genetic information.

Transforming Properties

The above principles, universally accepted today, were rejected outright not long ago. In fact, prior to the 1950s the general view was that nucleic acids were substances of somewhat limited cellular importance. The first convincing suggestion that DNA is the genetic material was made during the mid-1940s. The experiment involved the *transformation* of one type of pneumococcus, surrounded by the presence of a polysaccharide capsule and referred to as the S form because of its property of forming colonies with smooth-looking cellular perimeters, to a mutant without capsule, called the R form, which forms colonies with rough-looking outlines. These two forms are genetically distinct and cannot interconvert spontaneously. The transformation experiment demonstrated that a pure extract of DNA from the S form, when incorporated into the R form of pneumococcus, conveyed to the R form the specific property of synthesizing the characteristic polysaccharide capsule. Furthermore, the bacteria transformed from the R form to the S form maintained the property of synthesizing the capsule over succeeding generations. It was thus demonstrated that DNA was the *transforming* agent, as well as the material responsible for *transmitting* genetic information from one generation to the next. Almost three-quarters of a century had to elapse from the time nucleic acids were discovered until their important biological role was generally recognized.

DNA: a Molecule with Unusual Capacity

One of the striking characteristics of DNA is that it is able to encode an enormous quantity of biological information. An undifferentiated mammalian fetal cell contains only a few picograms (10^{-12} g) of

DNA. Yet this minute amount of material is sufficient to direct the synthesis of an enormous number of distinct proteins that will determine the form and biochemical behavior of a large variety of differentiated tissues in the adult animal.

The compactness with which such information is stored in DNA is unique. Even the sophisticated memory elements of contemporary computers would appear pitifully inadequate by comparison. How does DNA achieve such a supreme coding effectiveness? The answer must obviously be sought in the nature of its chemical structure. It turns out that this structure is not only consistent with the unique efficiency of DNA as a "memory bank," but also provides the basis for understanding how DNA eventually "translates" this information into proteins.

17.2 STRUCTURE OF DNA

Structurally DNA is a *polynucleotide*. A formal analogy between polynucleotides and proteins may therefore be perceived. Polynucleotides are the products of *nucleotide* condensation, just as proteins are produced by the polymerization of *amino acids*. As will become apparent later, this similarity of structures is an important element which facilitates the transfer of genetic information between these two distinct classes of macromolecules. The structure of nu-

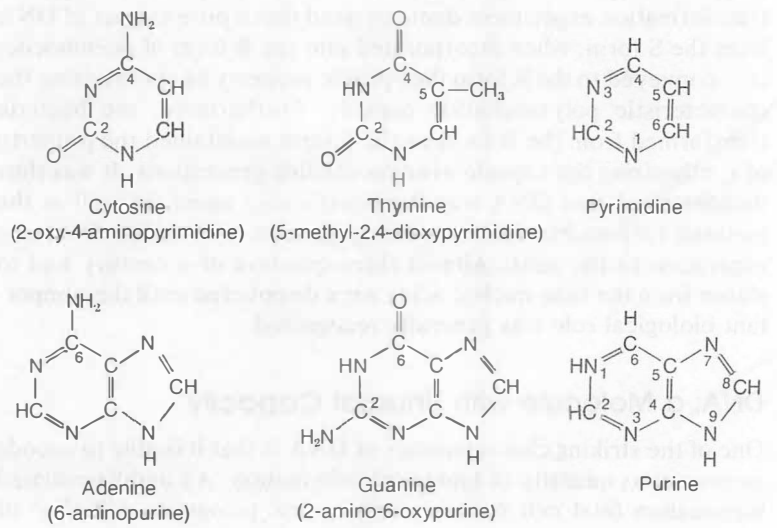


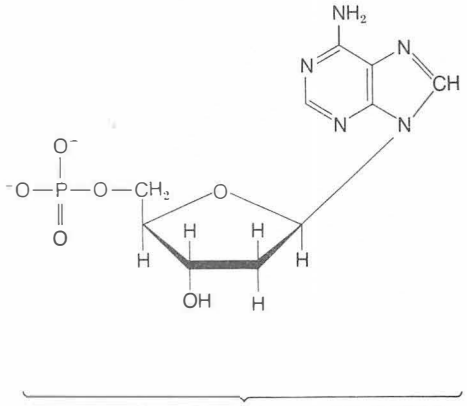
Figure 17.1
Structures of purines and pyrimidines.
Cytosine and thymine are related to the pyrimidine ring. Adenine and guanine are related to the purine ring system. Chemical names, rarely used in biochemistry, are based on the usual conventions and the numbering systems shown on the purine and pyrimidine rings.

cleotides and their constituent purine and pyrimidine bases are examined in Chapter 13 and reproduced in Figures 17.1 and 17.2.

The base composition of DNA varies considerably among species, particularly procaryotes, which have a range of 25–75% in adenine–thymine content. This range narrows with evolution, reaching limiting values of ~45–53% in mammals.

In addition to the four common bases, adenine, guanine, thymine, and cytosine, which occur in DNA from all sources, DNA isolated from many plant and animal tissues (e.g., wheat germ, thymus gland) contains small amounts of the base 5-methylcytosine. Methylated derivatives of the bases are also present in all DNA molecules examined to date. In addition, the DNA of certain bacteriophages (the T-even coliphages) contain 5-hydroxymethylcytosine in place of cytosine, and this derivative occurs in a glucosylated form. Even uracil, a base constituent of RNA, has been found in certain *Bacillus subtilis* phages, instead of thymine. The structures of some of these bases are shown in Figure 17.3.

The four major bases were first isolated at the turn of the century from the products of extensive hydrolysis of "nuclein." "Nuclein," a crude mixture of nucleic acids associated with proteins, was isolated from pus cells and salmon sperm cells during the latter part of the nineteenth century. The structures of the bases were determined from hydrolysates, using classic organic chemistry techniques. The structure of the 2-deoxyribose, however, escaped identification until the 1930s. By that time it had become apparent that the monomeric units of nucleic acids were nucleotides, but it was by no means clear how nucleotides were bonded together to form polynucleotide structures, DNA or RNA.



Deoxyadenosine-5′-monophosphate (αAMP)

Figure 17.2
Chemical structure of a deoxyribonucleotide. *Deoxyribonucleotides occur in the form of mono-, di-, and triesters of orthophosphoric acid. Nucleotides in the form of monophosphates, such as deoxyadenosine 5′-monophosphate, can be viewed as the monomeric units from which DNA polynucleotides are constructed. The other monomeric units are the 5′-monophosphates of the remaining three bases, cytosine, thymine, and guanine.*

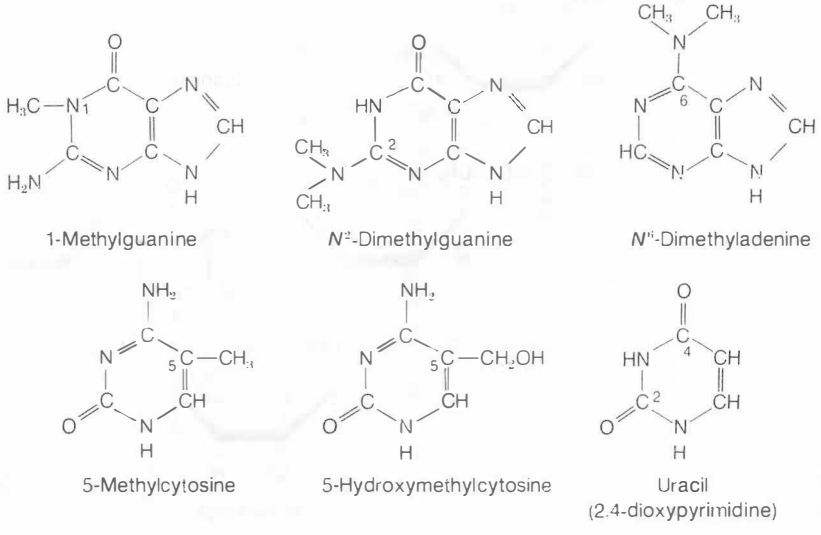


Figure 17.3
Structures of some less common bases occurring in DNA.

Polynucleotides

Polynucleotides are formed by the joining of nucleotides by phosphodiester bonds. The phosphodiester bond is the formal analog of the peptide bond in proteins. It serves to join, as a result of the esterification of two of the three hydroxyl groups of phosphoric acid, two adjoining nucleotide residues. Two free hydroxy groups are present in deoxyribose on the C-3′ and C-5′ atoms. Therefore these are the only hydroxyl groups that can participate in the formation of a phosphodiester bond. Indeed, it turns out that the nucleotide residues in DNA polynucleotides are joined together by 3′,5′-phosphodiester bonds, as shown in Figure 17.4.

In some instances polynucleotides are linear polymers. The last nucleotide residue at each of the opposite ends of the polynucleotide chain serve as the two terminals of the chain. It is apparent that these terminals are not structurally equivalent, since one of the nu-

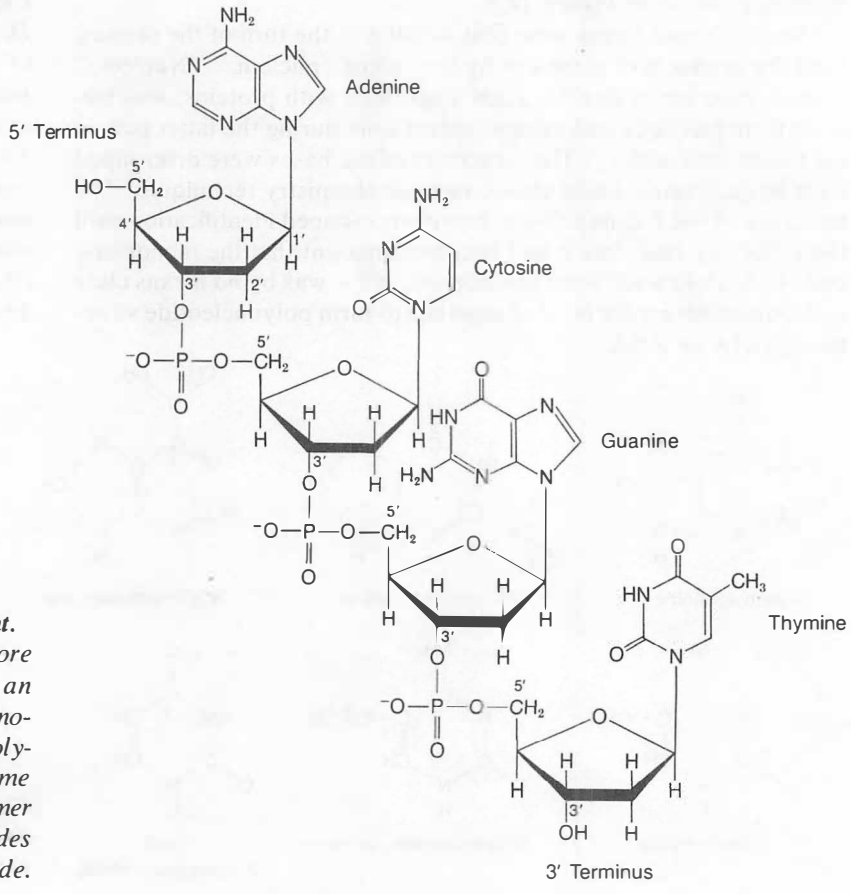


Figure 17.4
Structure of a DNA polynucleotide segment. *The example shown in this figure is, more precisely, a tetranucleotide, that is, an oligonucleotide consisting of four monomeric units. Although an exact cutoff polymerization size for this change in name does not exist, as a general rule a polymer containing less than 30 to 40 nucleotides is usually referred to as an oligonucleotide.*

cleotides must terminate at a 3′-hydroxyl group and the other at a 5′-hydroxyl group. These ends of the polynucleotides are referred to as the 3′ and the 5′ termini, and they may be viewed as corresponding to the amino and carboxyl termini in proteins. Although, in the example shown in Figure 17.4, the deoxyribosyl residues at both the 3′ and 5′ termini are present as free hydroxyl groups, it is not uncommon for cellular polynucleotides to have one or both of these hydroxyl terminals esterified. Polynucleotides also exist as cyclic structures, which contain no free terminals. Esterification between the 3′-OH terminus of a polynucleotide with its own 5′-phosphate terminus can produce a cyclic polynucleotide.

In this discussion long polymers of nucleotides joined by phosphodiester bonds are referred to as polynucleotides, in accordance with the prevailing nomenclature. A distinct name, oligonucleotide, is reserved for shorter nucleotide-containing polymers. According to formal rules of nomenclature, however, polynucleotides must be named by using roots derived from the names of the corresponding nucleotides, and using the ending *ylyl*. For example, the polynucleotide segment in Figure 17.4, in which the 5′ terminal is on the left of each nucleotide residue, should be named from left to right as

. . . deoxyaden*ylyl*, deoxycytid*ylyl*, deoxyguan*ylyl*, deoxythymid*ylyl* . . .

It is apparent, however, that the result of this approach is so cumbersome that abbreviations are generally preferred. For example, the oligonucleotide shown in Figure 17.4 is usually referred to as dAdCdGdT, and a polynucleotide containing only one kind of nucleotide, for example, dA, may be written as poly(dA). Oligo- and polynucleotide structures are also written out in shorthand, as shown in Figure 17.5. In every instance the sequence is written starting on the left with the nucleotide of the 5′ terminus.

DNA is made of polynucleotides, and it is the specific sequence of bases along a polynucleotide chain that determines the biological properties of the polymer. Although the structure of the nucleic acid building blocks, the bases, had been correctly known for many years, the polymeric structure initially proposed for DNA turned out to be one of the classical errors in the history of biochemistry. Experimental data obtained from what appears to have been partially degraded samples of DNA, and several other misconceptions, led to the erroneous conclusion that DNA consisted of repeating tetranucleotide units. Each tetranucleotide supposedly contained equimolar quantities of the four common bases. These impressions persisted to some degree until the late 1940s and early 1950s, when they were clearly shown to be in error. In the interim, however, these miscon-

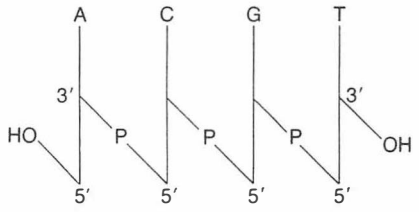


Figure 17.5
Shorthand form for structure of oligonucleotides.
The convention used in writing the structure of an oligo- or polynucleotide is a perpendicular bar representing the deoxyribose moiety, with the 5′-OH position of the sugar located at the bottom of the bar and the 3′-OH at a midway position. Bars joining the 3′ and 5′ positions represent the 3′,5′-phosphodiester bond, and the P on the left side of the perpendicular bar represents a 5′-phosphate ester. A 3′-phosphate ester is represented by placing the phosphate group on the right side of the bar. The base is indicated by its initial.

ceptions were responsible for setting back the acceptance of the concept that the DNA of chromosomes carried genetic information. The monotonous structure of repeating tetranucleotides appeared incapable of having the versatility to encode for the enormous number of messages necessary to convey hereditary traits. Instead proteins, which can be ordered in an almost unlimited number of amino acid sequences, were favored as the most suitable candidates for a hereditary function. The transformation experiment carried out in the mid-1940s, and the subsequent finding that DNA consists of polynucleotide rather than tetranucleotide chains, were responsible for the general acceptance of the hereditary role of DNA that followed.

Hydrolysis of the Phosphodiester Bond: Nucleases

The nature of the linkage between nucleotides to form polynucleotides was elucidated primarily by the use of exonucleases, which are enzymes that hydrolyze these polymers in a selective manner. *Exonucleases* cleave the last nucleotide residues in either of the two terminals of an oligonucleotide. Oligonucleotides can thus be degraded by the stepwise removal of individual nucleotides or small oligonucleotides from either the 5′ or the 3′ terminus. Nucleases sever the bonds in one of two nonequivalent positions indicated in Figure 17.6 as proximal (p) or distal (d) to the base, which occupies the 3′ position of the bond. For example, the treatment of an oligodeoxyribonucleotide with venom diesterase, an enzyme obtained from snake venom, yields deoxyribonucleoside 5′-phosphates. In contrast, treatment with a diesterase isolated from animal spleen produces deoxyribonuclease 3′-phosphates.

It should be noted that other nucleases, which cleave phosphodiester bonds located in the interior of polynucleotides and are designated as *endonucleases*, behave similarly in this respect. For instance, DNase I cleaves only p linkages, while DNase II cleaves d linkages. The points of cleavage along an oligonucleotide chain are indicated by arrows in Figure 17.6. Some *endonucleases* have been particularly useful in the development of early methodologies for sequencing of RNA polynucleotides. More recently other endonucleases, known as *restriction endonucleases*, have provided the basis for the development of recombinant DNA techniques.

Many nucleases do not exhibit any specificity with respect to base adjacent to the linkage that is hydrolyzed. Certain nucleases, however, act more discriminately next to specific types of bases or even specific individual bases. Restriction nucleases act only on sequences of bases specifically recognized by each restriction enzyme. Nucleases also exhibit specificities with respect to the overall structure of polynucleotides. For instance, some nucleases act on either single- or

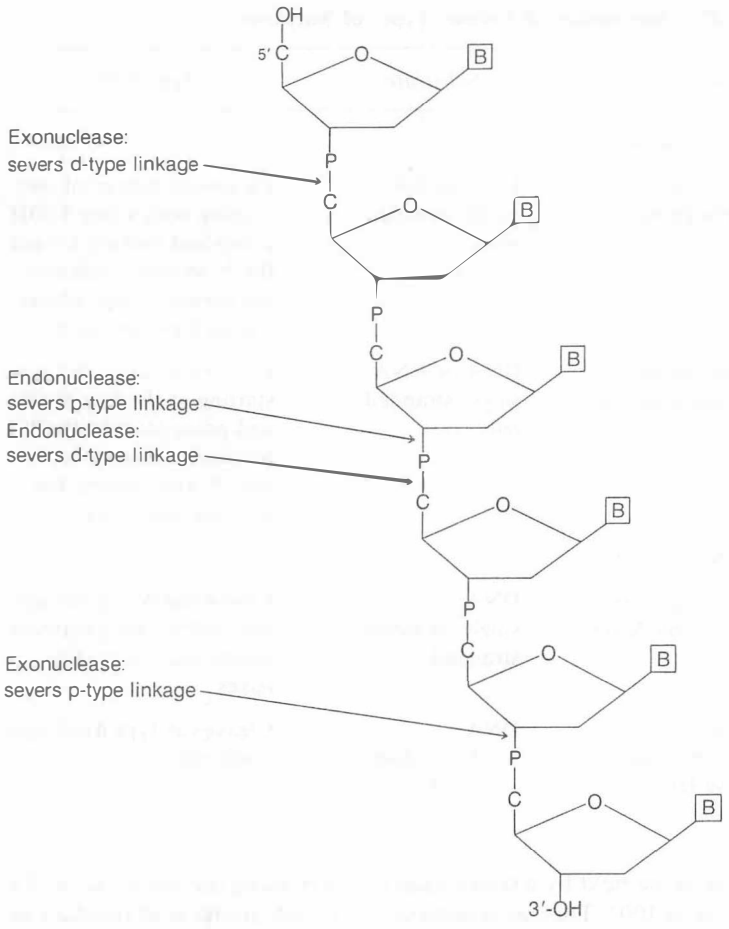


Figure 17.6
Nucleases of various specificities.
Exonucleases remove nucleotide residues from either of the terminals of a polynucleotide, depending on their specificity. Endonucleases hydrolyze interior phosphodiester bonds. Both endo- and exonuclease hydrolyze either d- or p-type linkages, as illustrated in the figure (see text for explanation of d- and p-type linkages).

double-stranded polynucleotides, whereas others discriminate between these two types of structures. In addition, some nucleases exclusively designated as *phosphodiesterases* will act on either DNA or RNA, whereas other nucleases will limit their activity to only one type of polynucleotide. The nucleases listed in Table 17.1 illustrate some of the diverse properties of these enzymes.

Secondary Structures of DNA

As has been emphasized previously, the polypeptide chains of protein are often arranged in space in a manner that leads to the formation of *periodic* structures. For instance, in the α helix each residue is

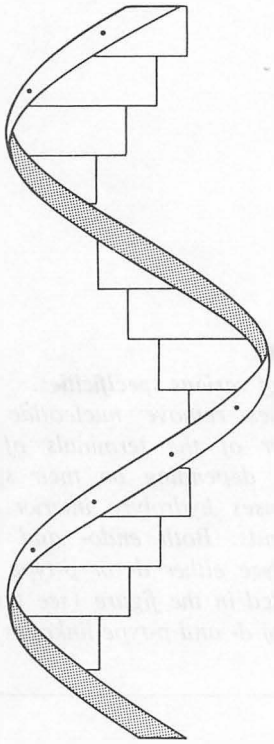


Figure 17.7
The conformation of a hypothetical, perfectly helical, single-stranded polynucleotide.
The helical band represents the phosphate backbone of the polynucleotide. The bases are shown in a side view as solid blocks in tight contact with their neighbors, above and below each base. The surfaces of the rings are in contact with each other and are not visible to the observer in the perspective from which the figure was drawn.

Table 17.1 Specificities of Various Types of Nucleases

Enzyme	*Substrate*	*Specificity*
EXONUCLEASES		
Snake venom phosphodiesterase	DNA or RNA single-stranded only	Cleaves all type p linkages, starting with a free 3′-OH group and moving toward the 5′ terminal; releases nucleoside 5′-phosphates; has no base specificity
Bovine spleen phosphodiesterase	DNA or RNA single-stranded only	Cleaves all type d linkages, starting at the free 5′-OH and proceeding to the 3′ terminal; releases nucleoside 3′-phosphates; has no base specificity
ENDONUCLEASES		
Bovine pancreas deoxyribonuclease (DNase I)	DNA single- or double-stranded	Cleaves all type p linkages but prefers those between purine and pyrimidine bases
Calf thymus deoxyribonuclease (DNase II)	DNA single- or double-stranded	Cleaves all type d linkages randomly

related to the next by a translation of 1.5 Å along the helix axis and a rotation of 100°. This arrangement places 3.6 amino acid residues in each complete turn of the polypeptide helix. The property of *periodicity* is also encountered with polynucleotides, which frequently occur in the form of helices.

Such preponderance of helical conformations among macromolecules is not surprising. The formation of helices tends to accommodate the effects of intramolecular forces, which in a helix can be distributed at regular intervals. The precise geometry of the polynucleotide helices varies, but the helical structure invariably results from the stacking of bases along the helix axis. In many instances stacking produces helices in which the bases are more or less perpendicularly oriented along the helix and touch one another. This arrangement, which obviously leaves no free space between two successive neighboring bases, is illustrated in Figure 17.7. Such stacked single-stranded helices, however, are not commonly encountered in nature. Rather, as it will become apparent from the

subsequent discussion, polynucleotide helices tend to associate with one another to form double helices.

Forces That Determine Polynucleotide Conformation

The hydrophobic properties of the bases are, to a large extent, responsible for forcing polynucleotides to adopt helical conformations. Examination of molecular models of the bases reveals that the edges of the rings contain polar groups (i.e., amino and hydroxyl group residues) that are able to interact with other polar groups or surrounding water molecules. The faces of the rings, however, are unable to participate in such interactions and tend to avoid any contact with water. Instead they tend to interact with one another, producing the stacked conformation. The stability of this arrangement is further reinforced by an interchange between the electrons that circulate in the π orbitals located above and below the plane of each ring.

Clearly then, single-stranded polynucleotide helices are stabilized by both *hydrophobic* as well as *stacking* interactions involving the π orbitals of the bases. The stability of the helical structures is also influenced by the potential repulsion among the charged phosphate residues of the polynucleotide backbone. These repulsive forces introduce a certain degree of rigidity to the structure of the polynucleotide. Under physiological conditions, that is, at neutral pH and relatively high concentrations of salts, the charges on the phosphate residues are partially shielded by the cations present, and the structure can be viewed as a fairly flexible coil. Under more extreme conditions the stacking of the bases is disrupted and the helix collapses. A collapsed helix is commonly described as a random coil. A conversion between a stacked helix and a random-coil conformation is depicted in Figure 17.8.

The DNA Double Helix

Although certain forms of cellular DNA exist as single-stranded structures, the most widespread DNA structure is the double helix. The double helix can be visualized as resulting from the interwinding around a common axis of two right-handed helical polynucleotide strands. The two strands achieve contact through hydrogen bonds, which are formed at the hydrophilic edges of their bases. These bonds extend between purine residues in one strand and pyrimidine residues in the other, so that the two types of resulting pairs are always adenine–thymine and guanine–cytosine. A direct consequence of these hydrogen-bonding specificities is that double-stranded DNA contains equal amounts of purines and pyrimidines. Examination of space-filling models clearly indicates the structural compatibility of these bases in forming linear hydrogen bonds.

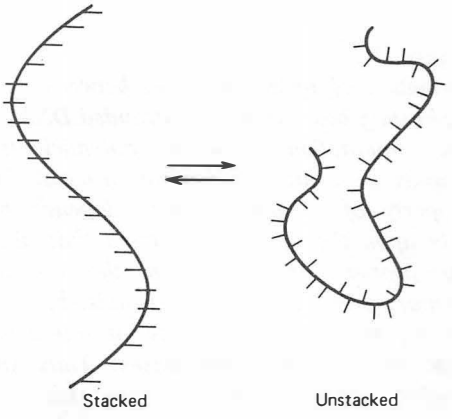


Figure 17.8
Stacked and unstacked conformations of a polynucleotide.
Stacking of the bases decreases the flexibility of a polynucleotide and tends to produce a more extended, often helical, structure.

This relationship between bases in the double helix is described as *complementarity*. The bases are complementary because every base of one strand is matched by a complementary hydrogen-bonding base on the other strand. For instance, for each adenine projecting toward the common axis of the double helix, a thymine must be projected from the opposite chain so as to fill exactly the space between the strands by hydrogen bonding with adenine. Neither cytosine nor guanine *fits precisely in the available space in a manner that allows the formation of hydrogen bonds across strands*. These hydrogen-bonding specificities, illustrated in Figure 17.9 ensure that the entire base sequence of one strand is complementary to that of the other strand.

The conventional double helix exists in various geometries designed as forms A, B, and C. These forms, however, share certain common characteristics. Specifically, the phosphate backbone is always located on the outside of the helix. Also, because the diesters of phosphoric acid are fully ionized at neutral pH, the exterior of the helix is negatively charged. The bases are well packed in the interior of the helix, where their faces are protected from contact with water. In this environment the strength of the hydrogen bonds that connect the bases can be maximized. The interwinding of the two polynucleotide strands produces a structure having two deep helical grooves that separate the winding phosphate backbone ridge.

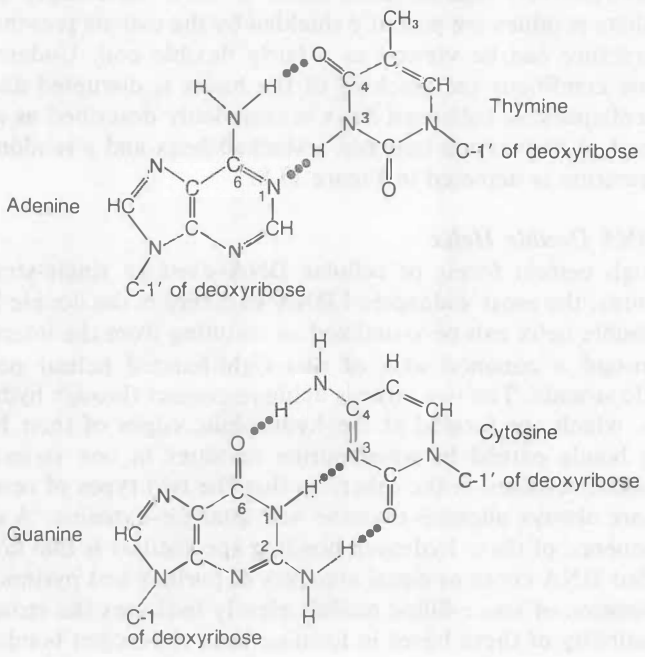


Figure 17.9
Formation of hydrogen bonds between complementary bases in double-stranded DNA.
The interaction between polynucleotide strands is a highly selective process. The property of complementarity depends not only upon the geometric factors that allow the proper fitting between the complementary bases of the two strands, but also on the electronic specificity of interaction between complementary bases. Thus specificity of interaction between purines and pyrimidines has also been noted both in solution and in the crystal form, and it is expressed in terms of strong hydrogen bonding between monomers of adenine and uracil or monomers of guanine and cytosine. In double-stranded DNA adenine interacts instead with thymine, which is a structural analog of uracil.

However, the precise geometry of the double helix varies among the different forms. The original x-ray data obtained with highly oriented DNA fibers suggested the occurrence of a form, later designated as B, which appears to be the one commonly found in solution and in vivo (Figure 17.10). A characteristic of this form is that one of its grooves is wider than the other, and it is referred to as the major groove to distinguish it from the second, or minor, groove. The nucleotide sequence of the polynucleotides can be discerned without dissociating the double helix by looking inside these grooves. As each of the four bases has its own orientation with respect to the rest of the helix, each base always shows the same atoms through the grooves. For instance, the C-6, N-7, and C-8 of the purine rings and the C-4, C-5, and C-6 of the pyrimidine rings line up in the major groove. The minor groove is paved with the C-2 and N-3 of the purine and the C-2 of the pyrimidine rings. Forms A and C differ from B in the pitch of the base pairs relative to the helix axis as well as in other geometric parameters of the double helix, as shown in Figure 17.11.

Recently obtained x-ray scattering patterns from DNA crystals suggest that DNA under certain circumstances also exists in other less conventional arrangements. These structures appear to consist of left- rather than right-handed strands, which exist in a "zigzag," rather than smooth, linear conformation. The precise nature, occurrence, and function of this type of DNA, which has been christened the Z form, is the subject of current investigation.

An important structural characteristic of double-stranded DNA is that its strands are *antiparallel.* Polynucleotides are asymmetric structures with an intrinsic sense of polarity built into them. As it may be concluded from inspection of Figure 17.12, the two strands are aligned in opposite directions that is, if *two adjacent bases in the same strand,* for example thymine and cytosine, are connected in the 5′ ⟶ 3′ direction, their complementary bases adenine and guanine will be linked in the 3′ ⟶ 5′ direction (directions are defined by linking the 3′ and 5′ positions within the same nucleotide). This antiparallel alignment produces a stable association between strands to the exclusion of the alternate parallel arrangement. Just as peptide geometries and the formation of α helices determine the overall preferred conformation of proteins, the formation of hydrogen bonds between complementary bases on antiparallel polynucleotide strands leads to the formation of the double helix.

The double-stranded structure for DNA was proposed in 1953. The proposal was partly based on the results of previously available x-ray diffraction studies, which suggested that the structures of DNAs from various sources exhibited remarkable similarities. These studies also suggested that DNA had a helical structure con-

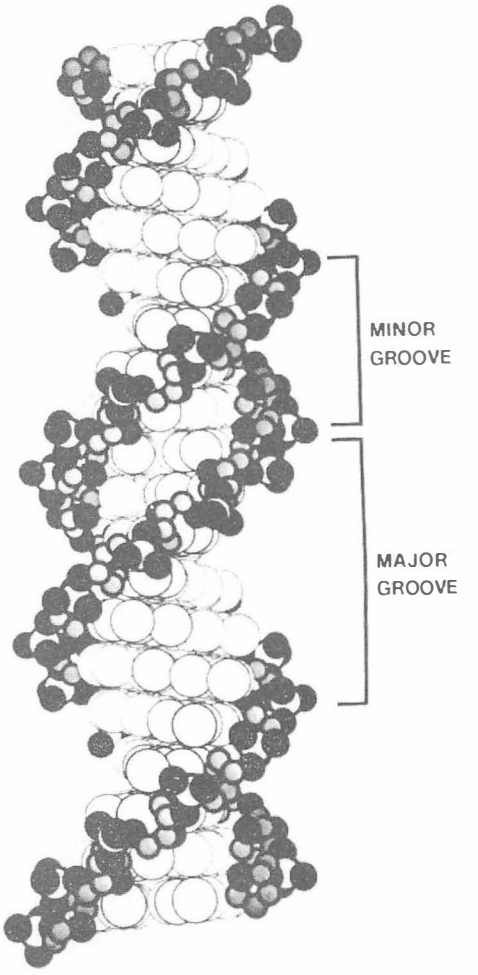


Figure 17.10
A space-filling model of the double helix.
Watson and Crick were the first to postulate a double-stranded model for the structure of DNA. The double helix is still referred to as the Watson and Crick model, although this structure has been substantially refined since it was proposed.
Reprinted with permission from R. Wells, *Prog. in nucleic acid research and molecular biology,* J. E. Davidson and W. E. Cohen, eds., New York Academic, Vol 24, 172, 1980. Copyright 1980, Academic Press.

taining two or more polynucleotides. An additional piece of evidence of central importance to the proposal was the clarification of the quantitative base composition of DNA, which was obtained independently in 1950. These results indicated the existence of molar equivalence between purines and pyrimidines, which turned out to be the essential observation suggesting the existence of complementarity between the two strands.

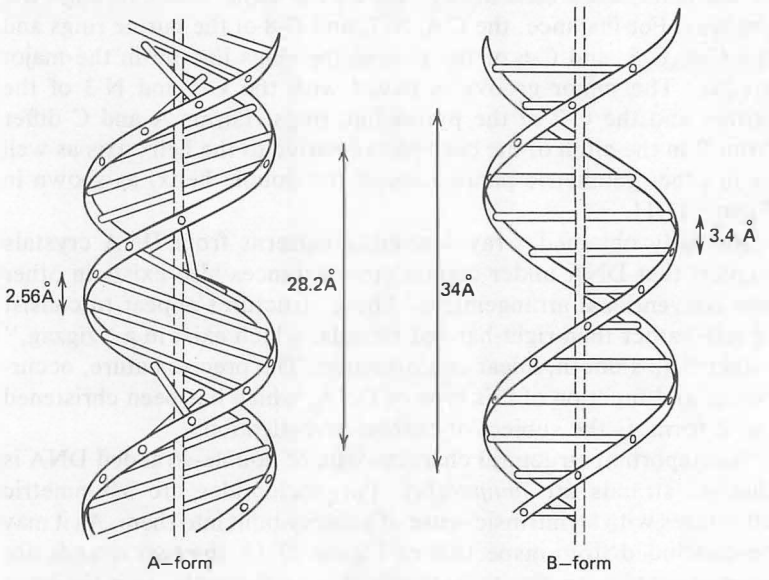


Figure 17.11
The various geometries of the DNA-double helix.
Depending on conditions, the double helix can acquire various forms of distinct geometries. In the B form of DNA the centers of the bases are ~3.4 Å apart and produce a complete turn of a helix with a pitch of 34 Å. Such an arrangement results in a complete turn of the helix for every 10 base pairs. The diameter of the helix is 20 Å. Form C (not shown) is very similar to the B structure, with a pitch of 33 Å and 9 base pairs per turn. Form A, which is obtained from form B when the relative humidity of the fiber is reduced to 75%, differs from B in that the base pairs are not perpendicular to the helical axis but are tilted. This tilt results in a pitch of 28.2 Å and a shortening of the helix by the packing of 11 pairs per helical turn.
From W. Guschelbauer, *Nucleic Acid Structure* Berlin: Springer-Verlag, 1976.

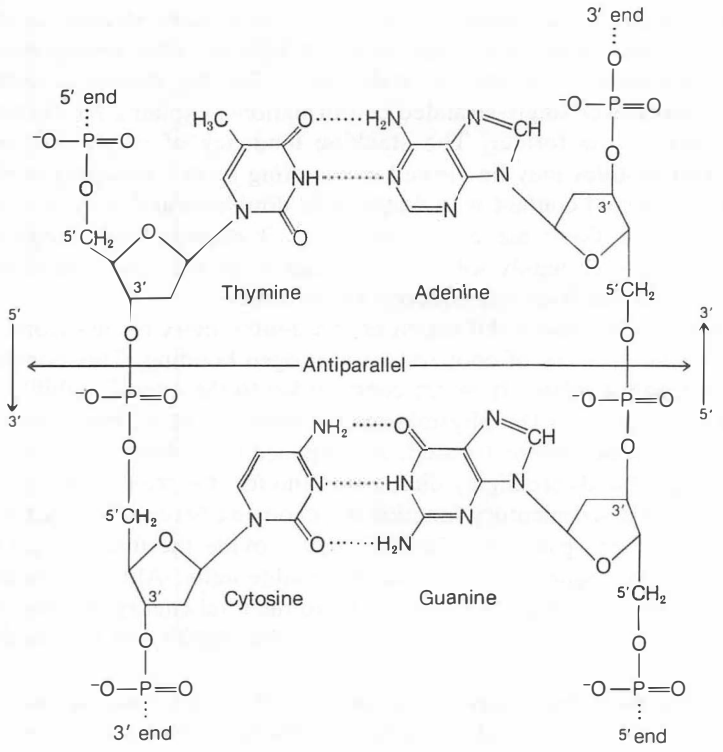


Figure 17.12
Antiparallel nature of the DNA strands.
The strands of a double-stranded DNA are arranged in such a manner that, as the complementary bases pair with one another, the two strands are aligned with opposite polarities, that is, the conventional assignment of the 5′ ⟶ 3′ direction to each of the strands suggests opposite directions. It should be noted that the geometry of the helices does not prevent a parallel alignment, but such an arrangement is not found in DNA.

It should be emphasized, however, that, at the time the double-stranded structure was proposed, no clear-cut evidence was available that DNA indeed had two strands. Since then several lines of evidence have confirmed the double-stranded nature of the helix. One of the most convincing experiments, based on the process of DNA replication, was carried out in the early 1960s. In this experiment one of the two DNA strands was made denser by the substitution of the heavier base analog 5-bromouracil for thymine. With this method the two complementary DNA strands were easily separated, fractionated by ultracentrifugation, and identified, establishing the double-stranded nature of the structure.

The Stability of the DNA Structure

The same factors that stabilize single-stranded polynucleotide helices, *hydrophobic* and *stacking* forces, are also instrumental in stabilizing the double helix. The separation between the hydrophobic core of the stacked bases and the hydrophilic exterior of

the charged sugar–phosphate groups is even more striking in the double helix than with single-stranded helices. This arrangement, which produces substantial stabilization for the double-stranded structures over single-stranded conformations, explains the preponderance of the former. The stacking tendency of single-stranded polynucleotides may be viewed as resulting from a tendency of the bases to avoid contact with water. The double-stranded helix is by far a more favorable arrangement, as it permits the phosphate backbone to be highly solvated by water while the bases are essentially removed from the aqueous environment.

Some additional stabilization of the double helix results from its extensive network of cooperative hydrogen bonding. This bonding makes only a relatively minor contribution to the overall stability of the structure, but the physiological importance of hydrogen bonds should not be underestimated. As opposed to hydrophobic forces, hydrogen bonds are highly directional and for this reason are able to provide a discriminatory function for choosing between correct and incorrect base pairs. In addition, they provide the final margin of free energy required to stabilize the double helix. Although hydrogen bonds make a minor contribution to the total energy of stabilization, their contribution is essential for the stability of the double helix.

In the past, the relative importance of hydrogen bonding and hydrophobic forces in stabilizing the double helix was not always appreciated. However, studies on the effect of various reagents on the stability of the double helix have suggested that the destabilizing effect of a reagent is not related to the ability of the reagent to break hydrogen bonds. Rather, the stability of the double helix is determined by the solubility of the free bases in the reagent, the stability decreasing as the solubility increases. Some of these findings, summarized in Table 17.2, emphasize the importance of hydrophobic forces in maintaining the structure of double-stranded DNA.

Ionic forces also have an effect on the stability and the conformation of the double helix. At physiological pH in solution, the mutual repulsion between the charged phosphate groups forces the chain into a relatively rigid rodlike configuration. In the presence of cations or other charged groups (the basic side chains of proteins) these repulsive forces are, to various degrees, screened out, and the flexibility of the double helix is partially restored.

Denaturation

The double helix is stabilized by ~1 kcal per base pair. Therefore a relatively minor perturbation can produce disruption in double-strandedness, provided that only a short section of the DNA is involved. As soon as the relatively few base pairs have separated, they

Table 17.2 Effects of Various Reagents on the Stability of the Double Helix[a]

Reagent	*Adenine Solubility* $\times 10^{-3}$ *(in 1 M reagent)*	*Molarity Producing 50% Denaturation*
Ethylurea	22.5	0.60
Propionamide	22.5	0.62
Ethanol	17.7	1.2
Urea	17.7	1.0
Methanol	15.9	3.5
Formamide	15.4	1.9

SOURCE: Data from L. Levine, J. Gordon, and W. P. Jencks, *Biochemistry*, 2:168, 1963.

[a] The destabilizing effect of the reagents listed below on the double helix is independent of the ability of these reagents to break hydrogen bonds. Rather, the destabilizing effect is determined by the solubility of adenine. Similar results would be expected if the solubility of the other bases were examined.

close up again and release free energy, and then the adjacent base pairs unwind. In this manner minor disruptions of double-strandedness can be propagated along the length of the double helix. Therefore, at any particular moment the large majority of the bases of the double helix remain hydrogen bonded, but all bases can pass through the single-stranded state, a few at a time. This dynamic state of the double helix is characterized by the movement of an "open-stranded" portion up and down the length of the helix, as indicated in Figure 17.13. The "dynamic" nature of this structure is an essential prerequisite for the biological function of DNA and especially the process of DNA synthesis.

Furthermore, the strands of DNA can be completely separated by increasing the temperature in solution. At relatively low temperatures a few base pairs will be disrupted, creating one or more "open-stranded bubbles." These "bubbles" form initially in sections that contain relatively higher proportions of adenine and thymine pairs. Adenine–thymine pairs are bound by two hydrogen bonds and are therefore less stable than guanine–cytosine pairs, which contain three such bonds per pair. As the temperature is raised, the size of the "bubbles" increases and eventually the thermal motion of the polynucleotides overcomes the forces that stabilize the double helix. This transformation is depicted in Figure

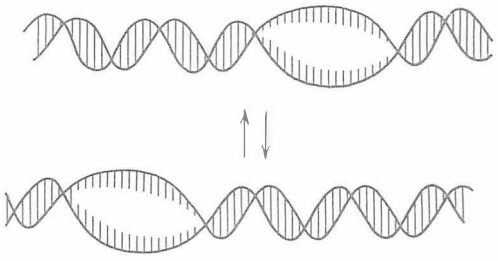


Figure 17.13
"Zipper" model for the DNA double helix. *DNA contains short sections of open-strandedness that can "move" up and down the helix.*

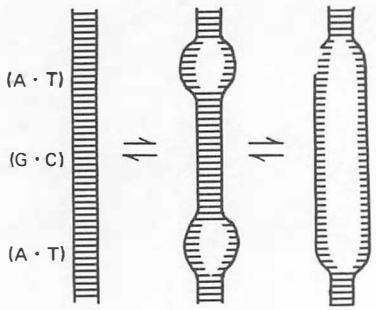


Figure 17.14
The structure of double-stranded DNA at increasing temperatures.
Disruptions of the double-stranded structure appear first in regions of relatively high adenine-thymine content. The size of these "bubbles" increases with increasing temperatures, leading to extensive disruptions in the structure of the double helix at elevated temperatures.

17.14. At even higher temperatures the strands can separate physically and acquire a random-coil conformation. This process, shown in Figure 17.15, is referred to as *denaturation.*

The process of denaturation is accompanied by a number of physical changes. Among these changes, buoyant density increases, viscosity is reduced, and ultraviolet absorbancy increases. Changes in absorbancy are frequently used for following experimentally the process of denaturation. DNA absorbs in the uv region due to the heterocyclic aromatic nature of its purine and pyrimidine constituents. Although each base has a unique absorption spectrum, all bases exhibit maxima at or near 260 nm. This property is responsible for the absorption of DNA at 260 nm. However, this absorbancy is almost 40% lower than that expected from adding up the absorb-

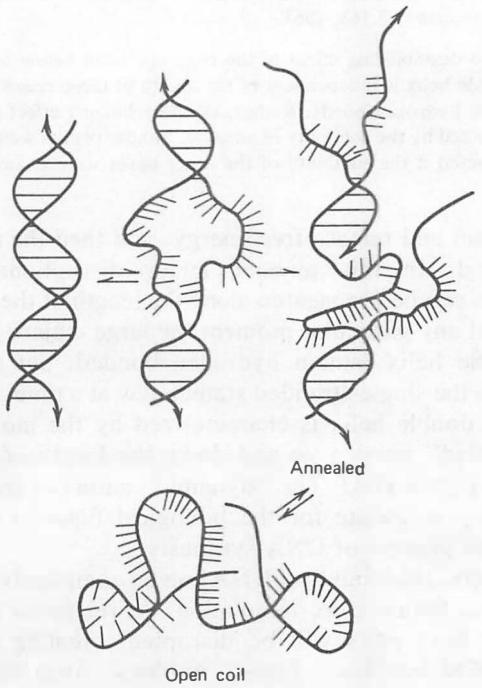


Figure 17.15
Denaturation of DNA.
At high temperatures the double-stranded structure of DNA is completely disrupted, with the eventual separation of the strands and the formation of single-stranded open coils. Denaturation also occurs at extreme pH ranges or at extreme ionic strengths.

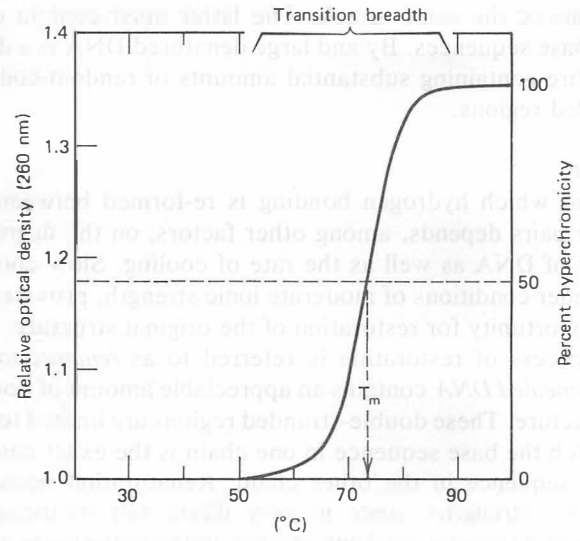


Figure 17.16
Temperature–optical density profile for DNA. *When DNA is heated, the optical density increases with rising temperature. A graph in which optical density versus temperature is plotted is called a "melting curve." Relative optical density is the ratio of the optical density at the temperature indicated to that at 25°C. The temperature at which one-half of the maximum optical density is reached is the midpoint temperature (T_m).* Redrawn from D. Friefelder, The DNA molecule: structure and properties. San Francisco: W. H. Freeman, 1978.

ancy of each of the base components of DNA. This property of DNA, referred to as *hypochromic effect,* results from the close stacking of the bases along the DNA helices. In this special arrangement interactions between the electrons of neighboring bases produce a decrease in absorbancy. However, as the ordered structure of the double helix is disrupted at increasing temperatures, stacking interactions are gradually decreased. Therefore, a totally disordered polynucleotide, a random coil, eventually approaches an absorbance not very different from the sum of the absorbancy of its purine and pyrimidine constituents.

Slow heating of double-stranded DNA in solution is accompanied by a gradual change in absorbancy as the strands separate. However, since the interactions between the two strands are cooperative, the transition from double-stranded to random-coil configuration occurs over a narrow range of temperatures, as indicated in Figure 17.16. The midpoint temperature, T_m, of this process, under standard conditions of concentration and ionic strength, is characteristic of the base content of each DNA. The higher the guanine-cytosine content, the higher the transition temperature between the double-stranded helix and the single strands. This difference in T_m's is attributed to the increased stability of guanine–cytosine pairs, as a result of the three hydrogen bonds that connect them in DNA, in contrast to only two hydrogen bonds that connect adenine and thymine pairs.

Rapid cooling of a heated DNA solution normally produces denatured DNA, a structure that results from the reformation of some hydrogen bonds either between the separate strands or between dif-

ferent sections of the same strand. The latter must contain complementary base sequences. By and large denatured DNA is a disordered structure containing substantial amounts of random-coil and single-stranded regions.

Renaturation

The extent to which hydrogen bonding is re-formed between the original base pairs depends, among other factors, on the degree of homogeneity of DNA as well as the rate of cooling. Slow cooling, especially under conditions of moderate ionic strength, provides the maximum opportunity for restoration of the original structure. This reversible process of restoration is referred to as *renaturation* or *annealing. Annealed DNA* contains an appreciable amount of double-stranded structure. These double-stranded regions are limited to sections for which the base sequence in one chain is the exact complement of the sequence in the other chain. Renaturation occurs at moderate ionic strengths, since in very dilute salt solutions the charged phosphate groups repel one another and keep the bases apart. Also, any reassociation of the two strands requires a minimum DNA concentration, below which renaturation is negligible.

The complementary strands of completely denatured DNA anneal by a two-step process. The first step involves the chance meeting of some complementary sequences on different strands, and it is a second-order reaction. The rate constant for this reaction is characteristic of each DNA, and it is inversely related to the size or, more precisely, to the number of the nucleotide pairs in DNA. This relationship can be observed in Figure 17.17 by comparing the rates of reassociation of MS-2, T4, and *E. coli* DNA which decrease as the size of the DNA genome increases. The first step of annealing is followed

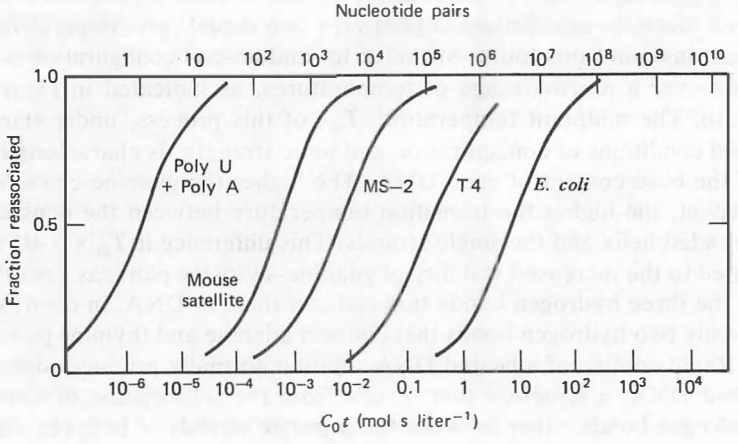


Figure 17.17
Reassociation kinetics for DNA isolated from various sources.
Each DNA is first fragmented to segments of approximately 400 nucleotides. The denatured segments are subsequently maintained in buffer at a temperature slightly below the T_m and are allowed to renature. The fraction of each polynucleotide reassociated, calculated from changes in hypochromicity, is plotted against the total concentration of nucleotides multiplied by the renaturation time (C_0t). The top scale shows the kinetic complexity of each DNA sample. Whenever a DNA contains repeated sequences, the repeated sequences are present in the fragments at higher concentrations than they would have been if a unique sequence had been fragmented. As a result, renaturation of fragments, obtained from DNAs containing repeated sequences, proceeds more rapidly the higher the degree of repetition. This is exemplified by the rates of renaturation of fragments obtained from the synthetic double-stranded polynucleotide poly(A)–poly(U) and mouse satellite DNA, a DNA that contains many repeated sequences. For a homogeneous DNA, which contains a distribution of different extents of repeated sequences, kinetic complexity can be defined as the minimum length of DNA needed to contain a whole single copy of the repeated sequence.
After R. J. Britten and D. E. Kohne, *Science*, 161:529, 1968.

by a first-order reaction, during which complementary base pairs located adjacent to the points of initial contact become rapidly hydrogen-bonded.

Hybridization

The self-association of complementary polynucleotide strands has also provided the basis for the development of the technique of *hybridization.* This technique depends on the association between any two polynucleotide chains, which may be of the same or different length, provided that a relationship of base complementarity exists between these chains. Hybridization can take place not only between DNA chains but also between appropriately related RNA chains as well as DNA–RNA combinations. In this section only DNA–DNA hybridizations are examined, but it should be apparent that the same principles apply to all types of hybridization.

Appropriate techniques have been developed for measuring the *maximum amount of polynucleotide* that can be hybridized as well as the *rates of hybridization.* The DNA to be tested for hybridization is first denatured. The resulting single strands are immobilized by binding to a suitable polymer, which is then used to pack a chromatography column. DNA grown in the presence of labeled precursors, usually tritiated thymidine, is allowed to run through the column that contains the bound, unlabeled DNA. The rate at which radioactivity is retained by the column obviously equals the rate of annealing between complementary strands.

Measurements of such rates have established that the DNA of eucaryotic cells contain a given nucleotide sequence reiterated a number of times. The principle of this determination is simple. For a DNA of a given size the rate of annealing depends on the frequency with which two complementary segments can collide with each other. Therefore, the larger the number of reiterated sequences in a given DNA, the greater is the chance that a particular collision will result in the formation of an annealed polynucleotide. On this basis the extent to which annealing takes place within a unit of time can be used to determine the number of reiterated sequences in the DNA.

Determinations of the maximum amount of DNA that can be hybridized have been used to establish homologies between the DNA of different species. This is possible because the base sequences of the DNA in each organism are unique for this organism. Therefore the annealed helices represent the same unique region of DNA even if the individual annealed strands originate from different cells. On this basis annealing can be used to compare the degree to which DNAs isolated from different species are related to one another. Consequently, the observed homologies serve as indexes of *evolutionary relatedness* and have been particularly useful for defining

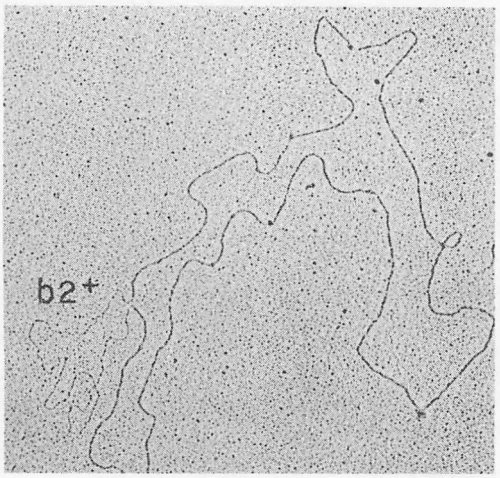


Figure 17.18
Heteroduplex formation in bacteriophage λ. *Electron micrograph of a heteroduplex DNA molecule constructed from complementary strands of bacteriophage λ and a bacteriophage λ deletion mutant (bacteriophage λβ2). In λβ2 a segment of DNA has been deleted, producing, at the site of deletion, a loop labeled b2+.*
Reprinted with permission from B. C. Westmoreland et al., *Science*, 163:1343, 1969. Copyrighted 1969 by the American Association for the Advancement of Science.

phylogenies in procaryotes. "Hybridization" studies between DNA and RNA have, in addition, provided very useful information about the biological role of DNA, particularly the mechanism of transcription.

Finally, the principle of *hybridization* has also served as the basis for the development of a technique that has permitted the construction of precise physical maps of DNA genes. This technique depends on the direct visualization under the electron microscope of single-stranded loops in the structures of artificially formed double-stranded DNA molecules known as *heteroduplexes*. The principle of this technique is simple. Heteroduplexes are constructed by hybridization of two complementary DNA strands. One of these strands, however, is selected on the basis that, as the result of a known mutation, it misses the gene being mapped. As is apparent from Figure 17.18, the complementary strands of the heteroduplex pair perfectly throughout the length of the molecule, with one important exception. Across from the position of the missing gene in the mutant strand the complementary strand forms a clearly visible loop. The position of the loop identifies the location of the deleted gene.

17.3 TYPES OF DNA STRUCTURE

The subject of DNA structure has been treated so far as though DNA were a "generic" substance, that is, only the essential features common to all DNAs have been presented. In fact, the specific structural features of DNA vary, depending on the origin and the function of each DNA molecule. DNAs differ in size, conformation, and topology.

Size of DNA

The size of DNA varies from a few thousand base pairs for the DNA of the small viruses, to millions for the chromosomal DNA of bacteria, and to billions for the chromosomal DNA of animals. Several types of expressions are commonly used to describe DNA size, including number of base pairs, molecular weights, the length of the strands, and even the actual weight of DNA. The units used in these expressions, however, can be easily interconverted, taking into account that a 1-million-mol-wt DNA contains approximately 1,500 base pairs which comprise a macromolecular segment of 0.5 nm

length. Also, since DNA is a macromolecule, DNA weight can be converted to molecular weight by division with the average molecular weight of a DNA nucleotide pair.

As is apparent from Table 17.3, the amount of DNA per cell increases as the complexity of the cellular function increases. It should be noted that although mammalian cells contain some of the highest amounts of DNA per cell, some amphibian, fish, and plant cells may contain even higher amounts. In fact, lung fish cells contain more than 40 times the amount of DNA in human cells, but such extraordinary amounts of DNA reflect a reiteration of nucleotide sequences within the DNA macromolecule and do not represent an actual increase in the size of DNA in terms of unique sequences. But aside from these minor irregularities, the size of the DNA of higher cells is very large indeed. The DNA contained within a single human cell, if it were stretched end to end, would be several meters long. This suggests that the polynucleotides are exquisitely packed in order to fit within the minute dimensions of the cell.

Because of their extraordinary length, relative to the total mass, the DNA molecules are extremely sensitive to shearing forces that develop during ordinary laboratory manipulations. Even careful pipetting may shear a DNA molecule. In addition, during the process of isolation it is difficult to prevent with absolute confidence the disruption of some phosphodiester bonds by contaminating endonucleases (nicking). For these reasons the precise size of DNA, especially that of the higher species, could not be determined until special handling techniques were developed, both for the isolation of DNA and the measurement of its molecular weight.

Techniques for Determining DNA Size

In any event, devising suitable methods for the measurement of the molecular size of DNA has been a scientific challenge. The classical methods for determining size in proteins, such as light scattering, sedimentation diffusion, sedimentation equilibrium, or osmometry proved to be unsuitable for measuring the molecular weight of even relatively small DNAs. For instance, because of the great mass of DNA the sedimentation coefficient of the macromolecule is so high that centrifuges could not be run slowly enough to yield useful data with existing methodology. Instead custom-tailored methods had to be devised. Equilibrium centrifugation in a density gradient (usually a concentrated cesium chloride solution), electron microscopy, and electrophoresis in agarose gels are among the principal methods providing reliable information about the molecular weights of various DNAs. Electron microscopy provides a measure of the length of DNA strands. Molecular weights can be calculated from known values of the mass per unit length. DNA can be visualized under the

Table 17.3 The DNA Cell Content of Some Species

Type of Cell	*Organism*	*DNA/Cell (pg)*[a]
Phage	T4	2.4×10^{-4}
Bacterium	*E. coli*	4.4×10^{-3}
Fungi	*N. crassa*	1.7×10^{-2}
Bird erythrocyte	Chicken	2.5
Mammal leukocyte	Human	3.4

SOURCE: B. Lewis. *Gene expression*, vol. 2, 2nd edition. New York: Wiley, 1980, p. 958.

[a] pg = picograms.

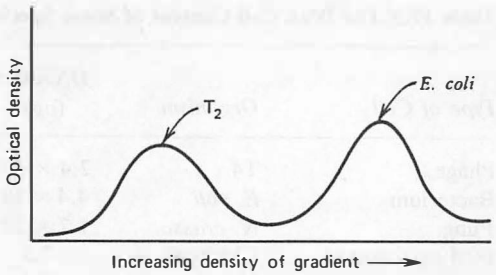


Figure 17.19
Equilibrium gradient centrifugation of DNA.
The DNA macromolecules travel into the increasingly dense regions of the gradient driven by centrifugal forces. The macromolecules equilibrate as soon as they reach an area of the gradient of density equal to their own. For example, bacteriophage T2 DNA and E. coli DNA can be resolved into two distinct bands. The width of the bands at equilibrium is related to the molecular weight of DNA.

electron microscope if it is first coated with protein and a metal film. Determination of molecular weights by electrophoresis depends on the molecular-sieving effect of porous agarose gels. Over a limited range of molecular weights the mobility of DNA is directly proportional to the logarithm of the molecule's weight. The range of the method is further extended by appropriate adjustments in the density of the agarose gels, which leads to changes in mobility.

In order to determine the molecular weight of DNA by equilibrium centrifugation a small portion of a DNA solution to be analyzed is layered on top of a gradient in a centrifuge tube. Upon centrifugation, the molecules of DNA sediment to equilibrium through the gradient. Under these conditions a homogeneous high molecular weight DNA will form a Gaussian band centered at a position in the gradient that corresponds to the density of the macromolecule. Molecules with different densities are resolved into a series of bands that sediment independently of one another, as shown in Figure 17.19. Interestingly, a relationship can be demonstrated between the width of the bands at equilibrium and the molecular weights. This relationship permits the determination of accurate molecular weights.

A biochemical method based on the labeling of the terminals of a macromolecule has been used successfully for determining molecular weights in proteins. A variation of this technique has been applied to DNA. In this case the macromolecule is treated with the enzyme alkaline phosphatase, which converts the 5′-phosphate nucleotide terminals of double-stranded DNA to the corresponding hydroxyl groups. These terminals are then esterified, using [γ-^{32}P]ATP with the enzyme polynucleotide kinase, which transfers ^{32}P from γ-labeled ATP. Thus the free 5′ terminus of each polynucleotide chain becomes labeled by the addition of a [^{32}P] phosphate residue, as shown in Figure 17.20. The labeled DNA is then analyzed by zonal centrifugation and detected from both its absorbancy at 260 nm and ^{32}P counting as indicated in Figure 17.21. The molecular weight is calculated from the ratio of the amount of ^{32}P to the absorbancy, both measured at the coinciding peaks of the bands. Clearly the

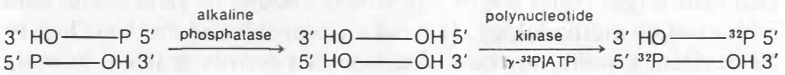


Figure 17.20
End-group labeling procedure.
The 5′ terminals on the opposite ends of DNA are labeled with ^{32}P by treatment with alkaline phosphatase and esterification of the resulting 5′-hydroxyl groups with [γ-^{32}P]ATP.

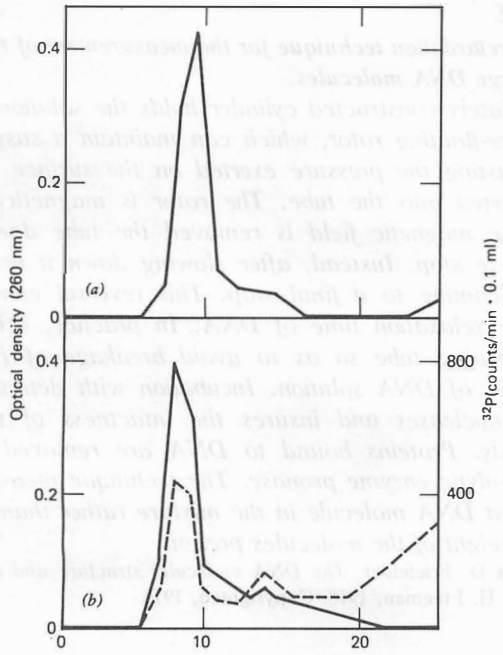


Figure 17.21
Zonal centrifugation profiles of denatured T7 DNA treated by the end-group labeling procedure.
Sedimentation is from right to left. (a) Untreated DNA. (b) DNA treated by the end-group labeling procedure. Zonal centrifugation is performed on a sucrose density gradient and should be distinguished from density gradient centrifugation. The latter is an equilibrium centrifugation with the macromolecules reaching equilibrium at regions within the tube at which their density equals the density of the environment. With zonal centrifugation the macromolecules move continuously until they reach the bottom of the tube or until the centrifuge is stopped. The molecular weight is calculated from the ratio of the amount of ^{32}P *(dotted line) to the optical density (solid line) at the peak of the curve.*
Redrawn from C. C. Richardson, *J. Mol. Biol.*, 15:49, 1966.

successful application of this technique requires homogeneous DNA from which possible fragments have been removed.

The above methods have permitted the determination of DNA molecular weights with an accuracy of at least 10% and often considerably better, but the usefulness of each method is limited within certain molecular weight ranges. Electrophoresis is most suitable for molecular weights in the range between 1.5×10^5 and 1.5×10^7. This range can be extended upward to 2×10^8 by electron microscopy. The most versatile method, however, is equilibrium centrifugation, the range of which extends approximately between 2×10^5 and 10^9. The high range of the method is limited because of the effect of shear forces on larger molecules. Therefore, because even bacterial DNAs often have molecular weights in excess of 10^9, it is apparent that none of the above methods can be used for very large DNAs. For DNA molecules of mol wt 10^{10} a specifically designed low shear viscometric method, described in Figure 17.22, has been developed. This method, known as viscoelastic retardation, is based on mildly stretching long DNA molecules by hydrodynamic shear forces. Once these forces are removed, the DNA molecules can relax back to their normal unstressed configuration. The relaxation time is related to, and can be used to determine with accuracy, the molecular weight of DNA molecules of the size found in eucaryotic chromosomes.

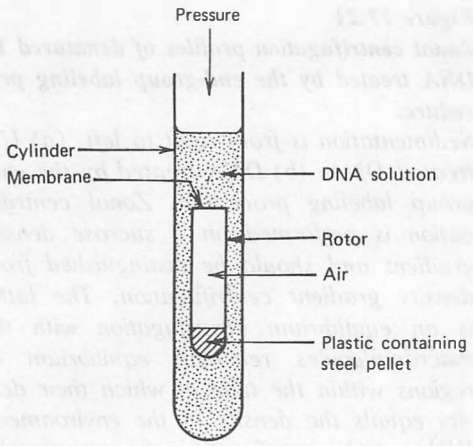


Figure 17.22
Viscoelastic retardation technique for the measurement of the molecular weight of large DNA molecules.
An appropriately constructed cylinder holds the solution to be measured. A free-floating rotor, which can maintain a suspended position by adjusting the pressure exerted on the surface of the solution, is inserted into the tube. The rotor is magnetically rotated, but once the magnetic field is removed the tube does not come to a complete stop. Instead, after slowing down it reverses direction before coming to a final stop. This reversal can be used to measure the relaxation time of DNA. In practice, cells are lysed in the measuring tube so as to avoid breakage of DNA caused by a transfer of DNA solution. Incubation with detergent at 65°C inactivates nucleases and insures the intactness of the resulting DNA strands. Proteins bound to DNA are removed by addition of the proteolytic enzyme pronase. The technique measures the size of the largest DNA molecule in the mixture rather than the average molecular weight of the molecules present.
Redrawn from D. Friefelder, *The DNA molecule: structure and properties.* San Francisco: W. H. Freeman, 1978. Copyrighted, 1978.

Linear and Circular DNA

The DNA of several small viruses occurs in the form of typical linear double-stranded helices of equal size. In addition, certain DNAs have naturally occurring interior single-stranded breaks. The breaks found in natural bacteriophage molecules result mostly from broken phosphodiester bonds, although occasionally a deoxyribose may be missing. The DNA of coliphage T5 consists of one intact strand and a complementary strand, which is really four different well-defined complementary fragments ordered perfectly along the intact strand. A similar regularity in the points of strand breaks is noted with a few other DNAs, for example, *Pseudomonas aeruginosa* phage B3, but generally interior breaks seem to be randomly distributed along the strands. The overall structure of the double helix is maintained because the breaks that occur in one strand are generally in different locations from breaks in the complementary strand.

Double-Stranded Circles

A very interesting aspect of the structure of DNA is that most naturally occurring DNA molecules exist in circular form. In some instances circular DNA exists even as interlocked circles. Provided that suitable precautions are taken to avoid shearing the DNA, the

circular form can be isolated intact and observed by electron microscopy. The circular structure may, in principle, be visualized as resulting from the circularization of a hypothetical linear DNA by formation of a phosphodiester bond between the 3′ and 5′ terminals of the linear polynucleotides.

The circular nature of DNA of the small phage ϕX174 was first suspected from studies that showed no polynucleotide ends were available for reactions with exonucleases. Sedimentation studies also revealed that endonuclease cleavage yielded one rather than two polynucleotides. These suspicions were later confirmed by observation with electron microscopy.

During the early 1960s, after workable methods for avoiding the shear of large molecules were developed, the circular nature of the DNA chromosome of *E. coli* was demonstrated by the use of autoradiography techniques. Soon it became apparent that many other DNAs (e.g., those of mitochondria, chloroplasts, bacterial plasmids, and mammalian viruses) also existed as closed circles. For instance, it was noted that fowlpox viral DNA could not be irreversibly denatured, an observation consistent with a circular structure. Obviously the strands of a circular DNA cannot be irreversibly separated because they exist as intertwined closed circles. The absence of 3′ or 5′ termini apparently provides an evolutionary advantage because it endows the circular DNA with complete resistance toward exonucleases, which act by hydrolyzing the phosphodiester bond of terminal nucleotides only. Thus circularity may be a protective mechanism against cellular exonucleases, which insures the longevity of DNA.

The DNA of some bacteriophages exists in a linear double-stranded form, which has the tendency to circularize when it enters the host cell. The linear DNA form of bacteriophage λ of *E. coli*, for instance, has single-stranded 5′ terminals of 20 nucleotides each. These terminals have complementary sequences, so that an *open circle* structure can be formed when the linear λ molecule acquires a circular shape, which allows the overlap of these complementary sequences. Subsequently, the enzyme *DNA ligase,* which forms phosphodiester bonds between properly aligned polynucleotides, joins the 3′- and 5′-terminal residues of each strand and transforms the DNA into a covalently *closed circle,* as illustrated in Figure 17.23.

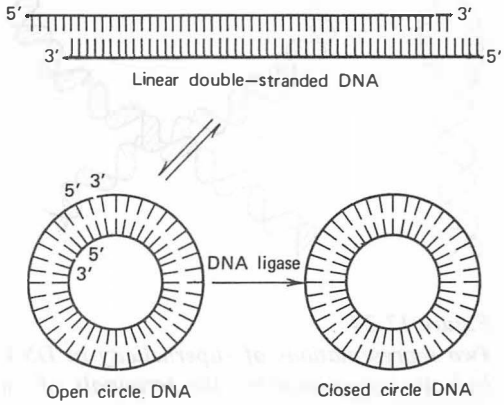


Figure 17.23
Circularization of λ DNA.
The DNA of bacteriophage λ exists in both a linear and a circular form, which are interconvertible. The circularization of λ DNA is possible because of the complementary nature of the single-stranded 5′ terminals of the linear form.

Single-Stranded DNA

With the exception of a few small bacteriophages (e.g., ϕX174, G4) that can acquire a single-stranded form, most circular as well as linear DNAs exist as double-stranded helices. The single-stranded nature of the nonreplicative form of ϕX174 DNA was first suspected in the 1950s when it was discovered that the base composition of

this DNA did not conform to the base *equivalance* rules, that is, for this DNA A ≠ T and G ≠ C. The single-stranded nature of this structure was also confirmed by the observation that the amino groups of the bases reacted rapidly with formaldehyde, which indicated that the bases were exposed. Furthermore, electron micrographs of ϕX174 indicated that single-stranded DNA appears more "kinky" and less thick than the double-stranded form. It may be noted that the discovery of the single-stranded circular form of ϕX174 actually preceded the identification of the replicative double-stranded form, which has a normal complementary base composition.

DNA Topology—Superhelices

The double-stranded circular DNAs, with few apparent exceptions, possess an intriguing topological characteristic. The circular structures contain fewer turns of the helix per unit length of the polynucleotide than does a linear double helix. These double helices, therefore characterized by a deficit of turns, are referred to as *negative superhelices.*

The thermodynamically favored structure of the double helix contains one complete turn per 10 base pairs. Therefore any deviation from this arrangement results in the formation of a strained structure. The deficit in turns characteristic of a superhelix can, in principle, be accommodated by the disruption of hydrogen bonding and the opening of the double helix over a small region of the macromolecular structure. The resulting structure may be viewed as consisting of a single-stranded loop along with a region of regularly spaced relaxed helical turns. Alternatively, the strain can be accommodated by the formation of tertiary structures with visible *supercoils* (Figure 17.24). These two representations of the negative superhelix should not be viewed as two distinct types of superhelicies, but rather as two manifestations of the same phenomenon. A distinctly different but related form of superhelix is obtained, however, if instead of a deficit, an excess of helical turns is present in the structure. In this case the strain in the resulting helix, which is referred to as a *positive superhelix,* can be accommodated only through the formation of supercoils.

The notion of superhelicity is often difficult for a beginner to grasp fully without examining an appropriate physical model. In the absence of a more suitable alternative, the unconvinced reader might attempt to twist two pieces of *fully extended* thick rope past the point at which considerable resistance develops. At that point the rope would represent a positive supercoil, which, in order to be accommodated without undue strain, must be allowed to escape the fully

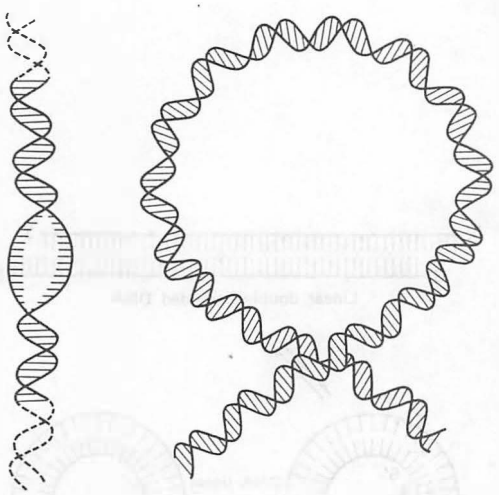


Figure 17.24
Two representations of superhelicity in DNA. *In both representations the terminals of the macromolecules must be viewed as unable to rotate freely around the axis of the double helix; if they did the structures would revert to the relaxed forms. Although the supercoiled structure on the right is a stable conformation, the structure of the left may be viewed as only a transient form. Both structures are characterized by a deficit of helical turns.*

extended conformation and acquire the form of a compact coil. This model also highlights the concept that superhelicity is inseparably associated with the existence of a closed or restricted topological domain. Superhelicity in this example will be preserved only for as long as both hands grasp the rope firmly so as to maintain a closed topological system. Once this closed system is interrupted, the superhelix can unwind and acquire a relaxed form.

Geometric Description of Superhelical DNA

The conformations acquired by the interlocking rings of a closed circular complex can be formally characterized by three parameters: the linking number α, the number of helical turns β, and the number of supercoils or tertiary turns τ. These parameters are related by the equation $\alpha = \beta + \tau$. The nature of β and τ is self-explanatory. When interlocked rings are viewed with one ring held in a plane, the linking number α may be defined as the number of times one ring passes through the other. As is apparent in Figure 17.25, α can also be determined by counting the number of times the two rings appear to cross each other and dividing this number by 2. This is because for each turn of the helix of a closed complex the second strand must pass through the circle formed by the first twice when viewed perpendicular to the helix axis.

Two important conclusions can be reached from consideration of these definitions and from examination of Figure 17.26. First, it is apparent that for every *relaxed* DNA the linking number and the number of helical turns are identical. However, as will be apparent shortly, the reverse is not true. Second, DNAs with a specific linking number can acquire various different arrangements in space. In the case of superhelical DNAs different types of supercoils may be formed. However, *all conformations with the same linking number α are interconvertible without the need of breaking any covalent bonds.*

The various forms of supercoiled DNAs can be described using the α, β, and τ numbers. The mental exercise shown in Figure 17.26 illustrates how these numbers apply. It should be recalled that the turns of the typical double helix are right-handed. Therefore, if a hypothetical linear DNA duplex that is 10 turns long ($\alpha = 10$ and $\beta = 10$) is unwound by, say, one turn, the resulting structure will have the following characteristics: $\alpha = 9$ and $\beta = 9$. A *potentially* equivalent structure can be formed if instead the ends of the same hypothetical DNA are secured so that they cannot rotate and the molecule is looped in a counterclockwise manner. Since in this case untwisting is not permitted to occur, the number of helical turns remains unchanged, that is, $\beta = 10$. However, as a result of the "looping" operations, the linking number is now reduced by 1, that is, $\alpha = 9$. The structure resulting from this deliberate introduction of

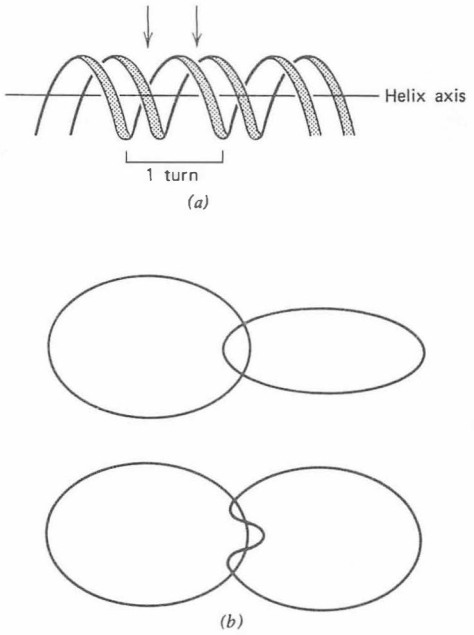


Figure 17.25
Determination of the linking number α in superhelical DNA.
(a) Side view of a schematic representation of the double helix. Note that the strands cross twice for each turn of the helix. (b) DNA circles interwound once and twice. Note that each pair of crossings is equivalent to one interwind.

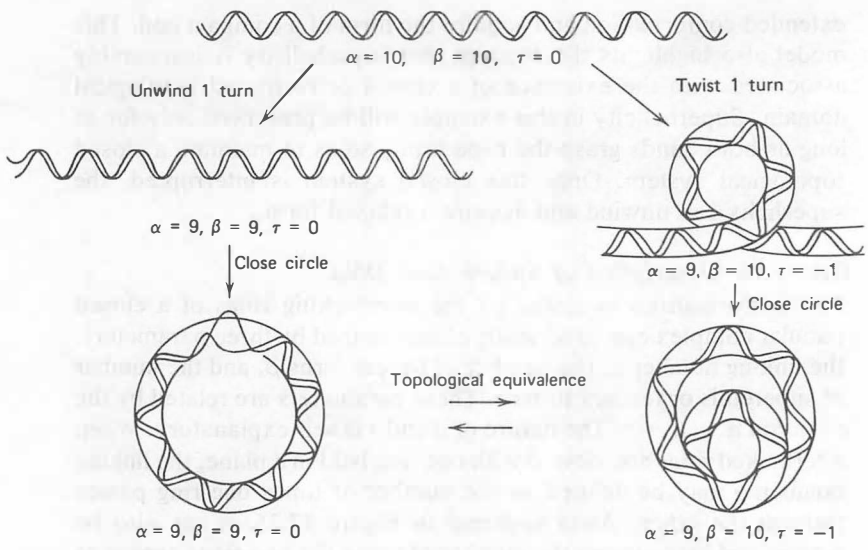


Figure 17.26
Various types of DNA superhelices.
An accurate representation of superhelical DNA structures can be made, using the number of helical turns β and the number of supercoils or tertiary turns τ along with a third parameter referred to as the linking number α as defined in the text. The figure shows ways of introducing one supercoil into a DNA segment of 10 duplex turns and the parameters of the resulting superhelices.
Redrawn with permission from C. R. Cantor and P. R. Schimmel, *Biophysical chemistry*, Part III, W. H. Freeman and Co., San Francisco, 1980. Copyrighted, 1980.

a loop is visibly superhelical. Furthermore, application of the equation that relates the values of α, β, and τ indicates that τ must be equal to -1, that is, the structure is a *negative* superhelix with *one* superhelical turn.

The two structures described above, $\alpha = 9$, $\beta = 9$, $\tau = 0$ and $\alpha = 9$, $\beta = 10$, $\tau = -1$ obviously have the same linking number and are therefore interconvertible without the disruption of any phosphodiester bonds. The potential equivalence of these two types of structure becomes more apparent when the ends of the polynucleotides in each structure are joined into a circle without the strands being allowed to rotate. Circularization produces an underwound circular structure and a doughnut-shaped superhelical arrangement referred to as a *toroidal* turn, which are freely interconvertible. A

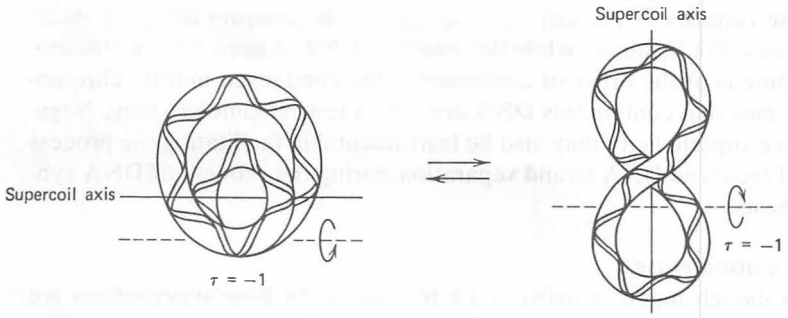


Figure 17.27
Equilibrium between two equivalent supercoiled forms of DNA.
The forms shown are freely interconvertible by unfolding the doughnut-shaped toroidal form along an axis parallel to the supercoil axis or by folding the 8-shaped interwound form along an axis perpendicular to the supercoil axis. The two forms have the same α, β, and τ numbers.
From C. R. Cantor and P. R. Schimmel, op. cit.

third equivalent structure, called an *interwound* turn, shown in Figure 17.27, can be produced by unfolding a toroidal turn along an axis which is distinct from the supercoil axis.

In summary, if the termini of a linear DNA molecule are covalently attached, a "relaxed" covalent circle results. However, if one end of the double helix is maintained in a fixed and stationary position while the other end is rotated in either direction prior to closing the circle, the resulting structure will twist in the opposite direction so as to generate a superhelical structure. For each additional complete turn of the helix, the circle will acquire one more superhelical twist in the opposite direction of the rotation in order to relieve the intensifying strain. As a result, topologically equivalent structures, such as those shown in Figure 17.26, will be created. A real superhelical DNA must exist as an equilibrium among these forms and many other intermediate arrangements in space that have the same linking number.

Although the closed circular form of DNA is an ideal candidate for acquiring a superhelical structure, any segment of double-stranded DNA that is in some way immobilized at both of its terminals qualifies for superhelicity. This property therefore is not the exclusive province of circular DNA. Rather, any appropriately anchored DNA molecule can acquire a superhelical conformation.

The DNA of animal cells, for instance, normally associated with nuclear proteins, falls into this category. Because of the fragility and the large size of this DNA, it has been difficult to establish whether it generally consists of a single circular piece, although this may be the case. However, even in the absence of a circular structure, animal DNA can acquire a superhelical form because its association with nuclear proteins creates numerous closed topological domains. In addition, most bacterial phages, animal viruses, bacterial plasmids, and cell organelles, such as mitochondria and chloroplasts, contain superhelical DNA. The existence of negative superhelicity appears to be an important factor, promoting the packaging of DNA within

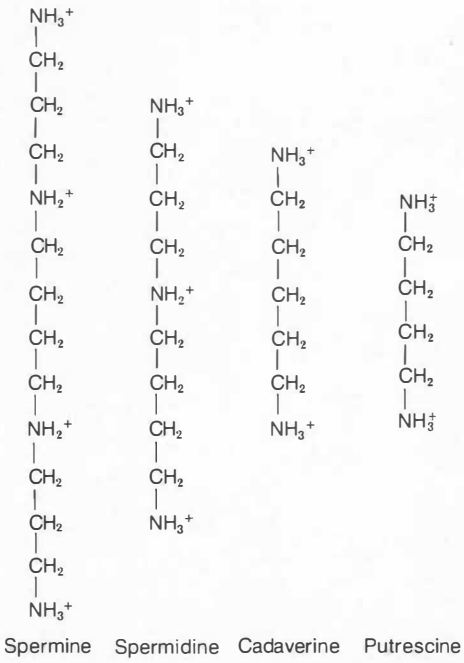


Figure 17.28
Structures of polyamines associated with procaryotic DNA.

the confines of the cell because supercoils generate compact structures. For instance, while the length of DNA in each human chromosome is of the order of centimeters, the condensed mitotic chromosomes that contain this DNA are only a few nanometers long. Negative superhelicity may also be instrumental in facilitating the process of localized DNA strand separation during the process of DNA synthesis.

Topoisomerases

Although much remains to be learned as to how superhelices are generated, specific enzymes known as *topoisomerases* appear to regulate the formation of superhelices. Several types of topoisomerases have been characterized to date, including the omega protein (ω protein), which has been isolated from *E. coli* and catalyzes the relaxation of negatively supercoiled DNA, and *gyrase,* also isolated from *E. coli,* which catalyzes the introduction of negative supercoils using ATP as an energy source. Gyrases may act by nicking a strand, which allows the removal of a few helical turns, and then sealing the nick so as to generate a DNA with a net deficit of turns.

Separation of superhelical DNA from the corresponding relaxed or linear forms can be achieved by gel electrophoresis or by equilibration centrifugation. With the latter method separation is achieved because the density of supercoiled DNA differs from that of the relaxed forms.

Nucleoproteins

As a rule DNA is associated in the cell with various types of protein. In *procaryotic cells* it is generally present as a double-stranded circular supercoil, which is, in part, associated with the inner side of the plasma membrane. Procaryotic DNA, however, does not complex with appreciable amounts of protein; instead the ionic charges of the phosphate residues are neutralized by various cations and polyamines such as spermine, spermidine, putrescine, and cadaverine (Figure 17.28). The small amounts of protein associated with procaryotic DNA appear to be essential for the process of protein synthesis rather than for the maintenance of electric neutrality. In contrast, almost all the DNA of differentiated eucaryotic cells is found in the nucleus as *chromatin.* There DNA is associated with equimolar amounts of highly basic polypeptides, the *histones.* In addition to histones, proteins of many other types are present in chromatin, and these proteins are grouped together under the somewhat unimaginative name of *nonhistone* proteins.

The histones, which as a result of their unusually high content of the basic amino acids lysine and arginine are highly polycationic in

nature, interact with the polyanionic phosphate backbone so as to produce uncharged nucleoproteins. In addition to contributing to the stabilization of the DNA structure through the formation of nucleoproteins, histones also are involved in the regulation of gene expression. For this reason there is little doubt that our understanding of the mechanism of gene regulation will increase as our knowledge of the structure of nucleoproteins becomes more exact.

Nucleohistone Structure

It is now established that the DNA–histone complex has a *periodic* structure. Histones are bound to DNA in clusters arranged at regular intervals that are separated from one another by *spacer* DNA regions consisting of approximately 50 base pairs each. Histones, regardless of their source, invariably consist of five types of polypeptides of different size and composition, as listed in Table 17.4. Each DNA–histone cluster consists of two molecules each of H2A, H2B, H3, and H4 histones, which are bound to a stretch of approximately 100–200 base pairs long. These clusters are referred to as nucleosomes or ν bodies (nu bodies). The clusters are formed by the wrap-

Table 17.4 The Structure of the Five Types of Histones[a]

Name	*Structure*[b]	*Residues*	*Molecular Weight*
H4		102	11,300
H3		135	15,300
H2A		129	14,000
H2B		125	13,800
H1		~216	~21,000

SOURCE: From D. E. Olins and A. L. Olins, *Am. Sci.*, 66:704, 1978.

[a] Histones, which are highly basic polypeptides, are often classified as lysine-rich (H1), slightly lysine-rich (H2A and H2B), and arginine-rich (H3 and H4). Many of the basic amino acids are clustered on amino-terminal tails (i.e., the first 30–40 amino acids on the N side). The nonbasic mid- and carboxyl-terminal portions of the histones (C side) form globular structures that appear to be the sites of interaction between histones in the nucleohistones. The basic tail, on the other hand, interacts with DNA. The H1 class is almost twice as large and more basic than the other histones. Its globular region is nearer the N terminal.

[b] Scale ⊢⊣ 10 amino acids.

ping of the DNA around the histone octamers in such a way as to form the supercoiled structure depicted in Figure 17.29. The remaining histone, H1, binds to DNA along the spacer regions.

The most "conserved" histones are the H4 and H3, which differ very little between extremely diverse species. The H2A and H2B histones are less highly conserved, but still exhibit substantial evolutionary stability, especially within their nonbasic portions. The H1 histones are quite distinct from the inner histones. They are larger, more basic, and by far the most tissue-specific and species-specific histones.

A small amount of eucaryotic DNA is located in the mitochondria and the chloroplasts of plant cells. This DNA, which occurs in the form of small superhelices, is generally free of protein.

Nucleosomes are universally present among eucaryotic organisms and appear to be the first level of chromosomal organization beyond the DNA helix. There is little doubt that the nucleosome has a definite packaging function for chromosomal DNA. Much remains to be

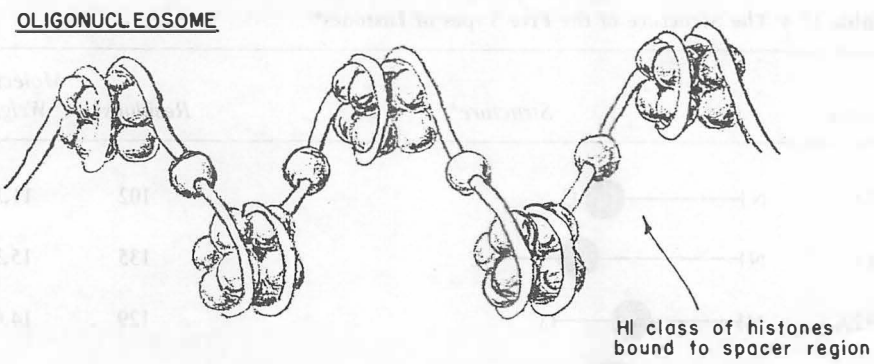


Figure 17.29

A model for the structure of nucleoproteins.

The DNA of eucaryotic cells is organized into a repeating arrangement of particles called nucleosomes. Nucleosomes are discrete, well-defined particles about 10 nm in diameter consisting of a closely packed octamer of histones, (H4, H3, H2A, and H2B)$_2$, surrounded by approximately 140 nucleotide pairs of DNA wrapped around the octamer in $1\frac{3}{4}$ shallow turns. Histone H1 is bound to spacer regions between nucleosomes. The zigzag organization of the oligonucleosomes can be seen on electron micrographs. The nucleosome appears to have a packaging function for DNA.

Redrawn with permission from D. E. Olins and A. L. Olins, Am. Sci., 66:704, 1978. Journal of Sigma Xi, The Scientific Research Society.

done before arriving at an understanding of the control of eucaryotic transcription and replication, but it is becoming increasingly apparent that both histone and nonhistone proteins are involved in these processes. While the dissociation of histones from chromosomal DNA might be a prerequisite for transcription, nonhistone proteins appear to provide more finely tuned transcription controls. However, the exact manner in which nonhistone proteins interact with DNA and regulate gene expression remains to be elucidated.

Finally, nonhistone proteins may also be involved in determining the structure of chromatin and may serve as sites for the binding of regulatory molecules.

Viral DNA is almost always complexed with protein. The function of the protein is generally one of "packaging." In essence the protein protects the DNA from mechanical damage or digestion by endonucleases by providing housing for the DNA within the tertiary structure of the protein. One unusual example of viral nucleoprotein, shown in Figure 17.30, is found in an animal virus, SV40. In this case, the protein, which is remarkably similar to the histones of eucaryotic chromosomes, forms a minichromosome characterized by repeating pockets of folded double-helical DNA and protein. Not surprisingly, once this protein is removed, the free viral DNA is found to acquire a superhelical conformation.

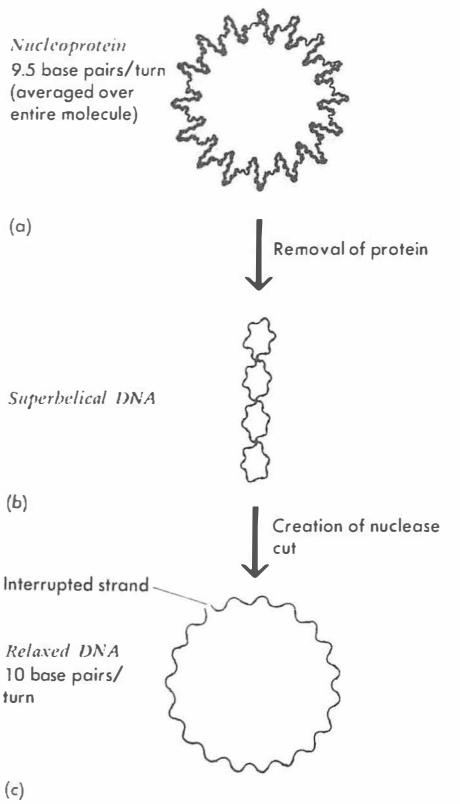


Figure 17.30
The nucleoprotein of virus SV40.
The organization of the nucleoprotein of virus SV40 exhibits similarities with that of eucaryotic nucleoprotein. The DNA of the virus is complexed with protein to form a "minichromosome" consisting of repeating units of folded double-helical DNA and protein. In this packing the DNA is superhelical and acquires the characteristic supercoiled conformation when the protein is removed.
Redrawn with permission from J. D. Watson, Molecular biology of the gene, 3rd ed., Menlo Park, Calif.: Benjamin-Cummings, 1977.

17.4 DNA STRUCTURE AND FUNCTION

Isolation of DNA

An obvious prerequisite for the study of the relationship between the structure and the function of DNA is the availability of isolated DNA in pure form. If the source of DNA is cellular rather than viral, DNA is normally extracted from a particular subcellular fraction of disrupted cells, such as nuclei and mitochondria. Cellular walls are usually disrupted by treatment with detergents, and DNA is precipitated by a mixture of ethanol and aqueous solution of sodium chloride. The DNA precipitate takes a fibrous form, and it is collected by winding the fibers on a glass rod. Protein and RNA form a nonfibrous precipitate, whereas most other cell constituents remain in solution. Because "spooled" DNA contains impurities, especially protein, it is shaken with water-saturated phenol, which produces a suspension of minute droplets of aqueous and organic phases. Protein precipitates at these aqueous–organic interphases, so that after

separation of the two phases protein remains at the interphase. DNA remains dissolved in the aqueous layer, from which it can be reprecipitated.

The main disadvantage of the method is that it produces DNA strands that are partially fragmented because of the hydrodynamic shear forces that develop during shaking. DNA is usually further purified by treatment with enzymes that degrade contaminant RNA and by chromatography on hydroxyapatite (calcium phosphate), which permits the separation of double-stranded DNA from single-stranded fragments.

Certain measurements requiring the presence of completely intact DNA can be carried out on cell lysates rather than isolated DNA. For instance, molecular weights can be measured on lysates obtained by treatment with detergents at high temperatures and in the presence of high salt concentrations and chelating agents. Under these conditions the freed DNA is not susceptible to partial hydrolysis by contaminating nucleases, and it is likely to remain completely intact.

Nucleotide Sequences in DNA

Overall *base composition* characterizes DNA only in a very general manner. Yet information on composition is often useful for DNA characterization. For determining this parameter, the nucleic acid is first hydrolyzed to its nucleotide components either by chemical or enzymatic means. The resulting nucleotides are separated usually by ion-exchange chromatography, and their amounts are determined spectrophotometrically. Alternatively, the composition can be indirectly estimated by equilibrium centrifugation of the DNA in a density gradient or by measurement of the melting temperature under standard conditions. In both of these techniques, the guanine-cytosine content of DNA influences in a quantitative manner the buoyant density and the thermal stability of the macromolecule, respectively.

A more specific property, which characterizes any DNA in a *unique* way, is its nucleotide *sequence*. Clearly the uniqueness of each DNA does not rest on its base composition but rather in the sequential arrangement of its individual bases. The direct determination of nucleotide sequences in DNA remained, until recently, an intimidating undertaking. This has been the case, in spite of the fact that the amino acid sequences of proteins and the nucleotide sequences of certain small RNAs such as tRNAs have been accessible for many years through the application of effective but tedious methods of digestion of these macromolecules by appropriate enzymes. These enzymes, which sever macromolecules at specific se-

quences, can be chosen so that they yield fragments with overlapping sequences from which overall sequences can be gathered. Such approaches could not be used for sequencing DNA, partly because even the smallest DNA molecules are very large compared to proteins or small RNAs. In addition, enzymes that cleave DNA next to either a specific base or a specific sequence were not available until recently.

Palindromes

The relatively recent discovery of *restriction endonucleases,* which cleave DNA chains in a specific sequence-dependent manner, has made possible the sectioning of large DNA molecules into small segments amenable to sequencing. These highly specific bacterial enzymes act in vivo by making double-stranded cuts in the DNA of an invading phage. This initial fragmentation exposes the DNA phage to eventual degradation by other nucleases. One important characteristic of the restriction enzymes is that they sever DNA only within, or in some instances near, specific nucleotide sequences that are four to six nucleotides long. These sites, known as *palindromes,* are characterized by a local symmetry, as illustrated by the examples listed in Table 17.5. The order of the bases is the same, or nearly

Table 17.5 Examples of Sites of Cleavage of DNA by Restriction Enzymes of Various Specificities[a]

Enzyme	*Microorganism*	*Specific Sequence*	*No. of Cleavage Sites for 2 Commonly Used Substrates*	
			ϕX174	pBR 322
EcoR1	*E. coli*	-G↓AATT-C- -C-TTAA↑G-	25	9
Hae III	*Haemophilus aegyptus*	-GG|CC- -CC↓GG-	11	22
Hpa II	*Haemophilus parainfluenzae*	-C↓CG-G- -G-GC↑C-	5	26
Hind III	*Haemophilus influenzae* Rd.	-A↓AGCT-T- -T-TCGA↑A-	0	1

[a] Cleavage takes place within palindromes. The cleavage sites are indicated by arrows.

the same, when the two strands of the palindrome are read in opposite directions. For example, in the case of the restriction enzyme EcoR1, isolated from *E. coli,* the order of the bases is GAATTC when read from the 5′ terminus of either of the strands.

The availability of restriction enzymes for sectioning large DNA sequences and the development of new techniques for separating DNA segments which differ from one another by only a single nucleotide has made the determination of sequences a simple matter. These sequencing techniques are described in Section 17.9. Also, the availability of sufficient amounts of isolated single-copy chromosomal genes for sequencing has been enormously facilitated by the development of recombinant DNA techniques, as described in Section 17.8.

Early attempts to determine DNA sequences were limited to small DNA regions, which could be easily sectioned off from the remaining DNA. Sequences that bind selectively with various functional proteins, for example, RNA polymerase and the repressor proteins, have been among the first to be determined. As a rule these sections can be separated from the remaining DNA by nuclease digestion of the complexes formed between DNA and the respective proteins. The protein protects the DNA section over which it is bound from the action of nuclease, and the protected DNA is recovered after digestion by dissociation of the protein. These studies indicated that

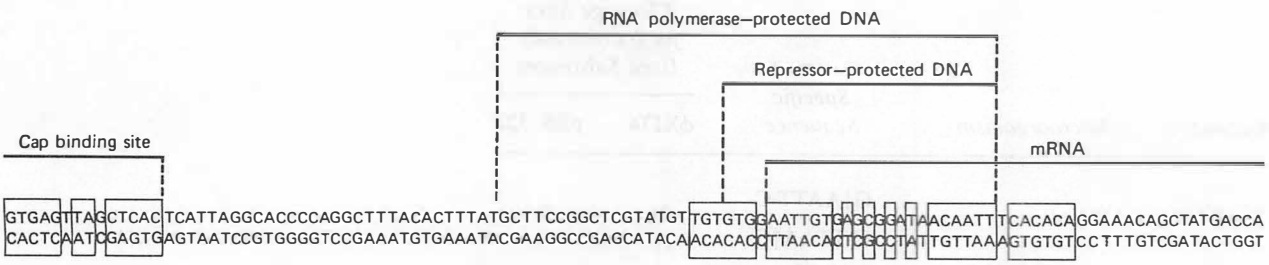


Figure 17.31

The nucleotide sequence of part of the DNA segment that controls the synthesis of the enzyme β-galactosidase in E. coli (the lac operon). *The binding regions of the cap protein, which acts as an activator of transcription, and of the lac repressor protein, an inhibitor of transcription, are indicated. Also shown is the region of RNA polymerase interaction. The presence of two palindromic sequences is indicated by boxes.*

From C. R. Cantor and P. R. Schimmel, op. cit. Part I.

many functional proteins and enzymes interact with DNA over regions of palindromic sequence (Figure 17.31).

Palindromes also serve as recognition sites for *methylases*, enzymes that modify the host DNA by introducing methyl groups into two bases of the palindrome. Once methylated, these palindromes cannot be recognized by the corresponding restriction enzymes, and the DNA of the host is protected from cleavage.

The new sequencing methods have made possible the determination of the complete nucleotide sequences of the DNA of many small viruses containing thousands of nucleotide residues, including ϕX174, G4, and SV40. The effectiveness of the new methods is such that sequencing the DNA of even higher cells is now becoming a routine undertaking.

Procaryotic DNA

The primary structure of small procaryotic DNAs reveals that structural genes, nucleotide sequences coding for protein, do not always have distinct physical locations in DNA. Rather they frequently overlap with one another, as illustrated by the partial sequence of bacteriophage ϕX174 shown in Figure 17.32. It is believed that this type of overlap provides for the efficient and economic utilization of the limited DNA present in small procaryotes. This arrangement of genes may also be a factor in controlling the sequence in which genes are expressed.

Eucaryotic DNA

Intervening Sequences. The case of eucaryotes used to be somewhat of a puzzle. Their DNA content appeared to be much too high, on the assumption that it consisted of only structural genes along with some sequences used to control gene expression, as is the case with procaryotes. Determination of the complete nucleotide sequences of whole eucaryotic DNA could have shed some light on this issue. Yet this would have been, and continues to be, an impractical task. The complete sequencing of large sections of eucaryotic DNA, for instance, structural genes and their surrounding regions, has now become a relatively easy undertaking. As a result, sequence data of eucaryotic DNA segments are accumulating rapidly. These data indicate that, in contrast to procaryotes, eucaryotic genes not only do not overlap, but with few exceptions (e.g., the genes of histones) are interrupted by intervening nucleotide sequences, as shown in Figure 17.33.

As a rule the sequence and the size of intervening sequences varies greatly among species, but generally intervening segments may be 5 to 10 times longer than the sum of the length of the parts of the

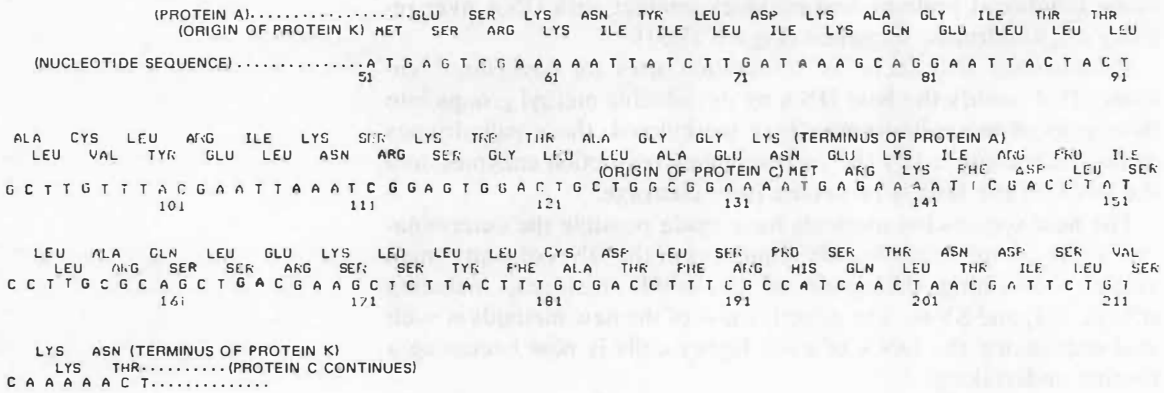


Figure 17.32

Partial nucleotide sequences of contiguous and overlapping genes of bacteriophage ϕX174.

The complete nucleotide sequence of ϕX174 is known. Only the sequence starting with nucleotide 51 and continuing to nucleotide 219 is shown in this figure. This sequence codes for the complete amino acid sequence of one of the proteins of ϕX174, protein K. A part of the same sequence, nucleotide 51 to nucleotide 133, codes for part of the nucleotide sequence of another protein, protein A (the remaining part of protein A is coded by a sequence, not shown, extending on the left beyond nucleotide 51). The remaining part of the sequence coding for protein K, which starts with nucleotide 133, also codes for part of a third protein, protein C. Similar overlaps are noted between other genes of ϕX174; for instance the sequence coding for a fourth protein, protein B, extends on the left of nucleotide 51.

Adapted with permission from M. Smith, *Am. Sci.*, 67:61, 1979. Journal of Sigma Xi, The Scientific Research Society.

structural genes they separate. Some genes are interrupted only once, whereas others are highly fragmented. For instance, the conalbumin gene of the chicken, which codes for a major protein in egg white, may be divided by intervening sequences in as many as 17 distinct sections.

At the present time only educated guesses can be offered regarding the biological function of intervening sequences. The possibility exists that these sequences play a role in the control of gene expression. In addition, the suggestion has been made that genes separated

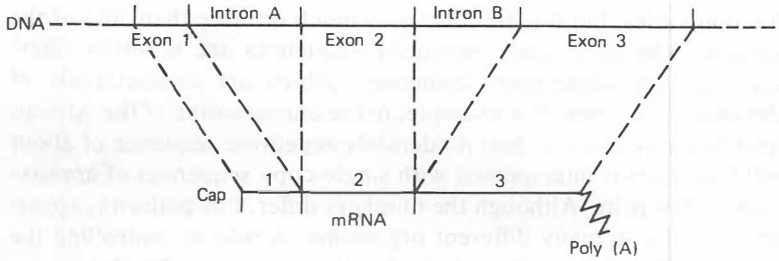


Figure 17.33
Schematic presentation of a eucaryotic gene. *The nucleotide sequences of eucaryotic genes are frequently separated by polynucleotide segments that are not transcribed to mRNA and therefore are not translated to protein. These segments are referred to as introns. The gene is thus separated into noncontiguous segments called exons. The top horizontal line in the figure represents a part of the DNA genome of a eucaryote; the bottom line the mRNA produced by it. In this hypothetical example the DNA consists of two introns and three exons. The intron sequences are not transcribed into RNA.*
From F. Crick, *Science*, 204:264, 1979. Copyrighted 1979 by the American Association for the Advancement of Science.

by intervening sequences could, on an evolutionary scale of time, be more easily shuffled to produce new gene combinations than genes that are put together as one piece.

Repeated Sequences. Until recently, the nucleotide sequences of eucaryotic DNA had been extensively studied by hybridization techniques. The more recent application of direct sequencing methods on DNA fragments, obtained by restriction endonuclease digestion, further extended the scope of these studies. As a result, in addition to obtaining the sequences of specific DNA sections, a good understanding of the complex characteristics of the primary structures of eucaryotic DNAs is now beginning to evolve. As distinct from procaryotes, in which all the nucleotide sequences in the chromosome occurs only once per haploid genome, the DNA of eucaryotes contains multiple copies of certain nucleotide sequences, which are repeated anywhere from a few times to millions of times per genome. Based on the distribution of repeat units, three classes of sequences have been distinguished in eucaryotes; the *nonrepetitive,* the *moderately repetitive,* and the *highly repetitive*. The nonrepetitive regions consist of unique sequences generally existing as single copies. These regions provide the instructions for the amino acid sequences of proteins, that is, they are the structural genes. Typically, a structural gene consists of a few thousand nucleotide pairs out of the several million pairs present in the total DNA. In some instances structural genes may be present as two or more identical or closely related copies. The obvious evolutionary advantage of this slight redundancy is that if a harmful mutation occurs in one of these two copies the synthesis of the corresponding protein will not be altogether prevented. Aside from this possibility, however, structural genes consist of unique sequences, which, as a rule, do not contain any elements of nucleotide sequence repetition.

The moderately repetitive fraction consists of copies of identical or closely related sequences that are repeated many times over, generally with a frequency of 10^3 to 10^6. The genome content in this

fraction varies, but it is frequently as much or more than 50% of the genome. The moderately repetitive sequences are normally interspersed with single-copy sequences, which are characteristic of the structural genes. For example, in the chromosome of the African toad *Xenopus laevis* a short moderately repetitive sequence of about 300 base pairs is interspersed with single-copy sequences of approximately 800 pairs. Although the numbers differ, this pattern is apparently similar in many different organisms. A role in controlling the transcription of the adjacent single-copy sequences has been proposed for these interspersed moderately repetitive sequences, especially those occurring with relatively low frequencies. This hypothesis is consistent with the surprising observation that genes not expressed in a tissue are nevertheless transcribed; and so are the interspersed moderately repetitive sequences. In fact, what appears to vary from tissue to tissue is not the extent to which genes are transcribed but rather the *extent to which the transcripts of the reiterated sequences are produced.*

In some instances moderately repetitive sequences may also exist as segregated tandem arrays. The two distinct types of arrangement of the moderately repetitive sequences, tandem and interspersed, appear to relate to different functions for these sequences. The tandem arrays have distinct properties, in that they are used for the synthesis of ribosomal and transfer RNA and for certain proteins of specialized function. For example, in some species of sea urchins four closely linked genes that code for histones are repeated several hundred times in a tandem array. Such arrays also include nucleotide sequences, which are not transcribed and which are referred to as spacer regions.

The arrangement of structural genes, with their intervening sequences, and the moderately repetitive segments of DNA of the interspersed and tandem array types is illustrated in Figure 17.34. In this hypothetical segment of eucaryotic DNA the single-copy and moderately repetitive sequences are indicated, which together normally account for more than 90% of the total nucleotide content of the genome.

A small part of the genome, usually 5–10%, consists of highly repetitive sequences. In most cases studied the repeating unit, which may be reiterated for as many as several million times, has been found to be typically less than 20 nucleotides long. More recently, however, some considerably longer repeat units have been identified. For instance, in the genome of the African green monkey a 172-base pair segment has been found to be highly repeated. Determination of the nucleotide sequence of the highly repetitive DNA reveals that there are few sequence repetitions within the 172-base segment. The reiterated units consist of a set of closely related but

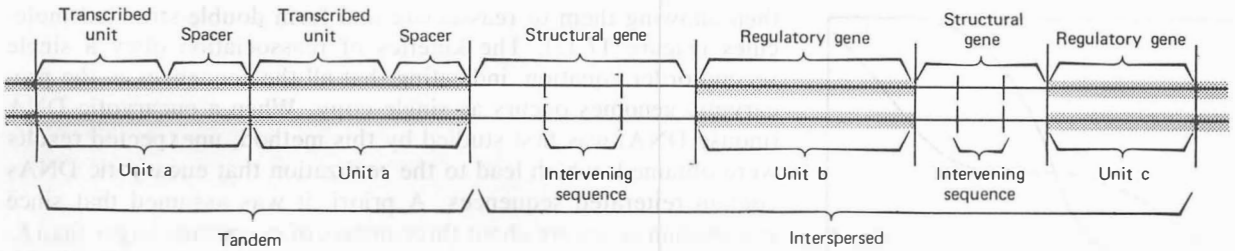


Figure 17.34

Structural genes with intervening sequences of the "interspersed" and "tandem array" types.

A hypothetical segment of eucaryotic DNA may be visualized as consisting of nonrepetitive sequences (indicated by a thin line) as well as moderately repetitive sequences (indicated by a shaded line). The latter can be of the interspersed or the tandem type. The interspersed sequences are separated by structural genes. The structural genes themselves are frequently interrupted by intervening sequences. The terms unit a, unit b, and unit c refer to distinct nucleotide sequences.

variant sequences.

Some of the highly repeated sequences have a distinct base composition from that of the remaining DNA. These sequences, which typically are rich in G–C content and exhibit higher buoyant densities, can be isolated from the total genomic DNA by shearing DNA into segments of a few hundred nucleotides each and separating the fragments by density gradient centrifugation. They are termed *satellite* DNA because after centrifugation they appear as satellites of the band of bulk DNA. For example, DNA of a satellite band from the kangaroo rat has a simple structure, consisting of the reiterated sequence 5′-GGACACAGCG-3′. Other highly repeated sequences cannot be isolated by density, but they can be obtained by virtue of their property of rapid reannealing. Some of the highly repetitive sequences can also be isolated by digestion of total DNA with restriction endonucleases that cleave at specific sites within the repeated sequence. The function of highly repetitive DNA is not known. However, since these reiterated sequences do not serve as templates for RNA synthesis, they are suspected to have a structural role, possibly in pairing and segregation of chromosomes.

It may be recalled that the genome size of procaryotic DNA can be determined by fragmenting the DNA, denaturing the fragments and

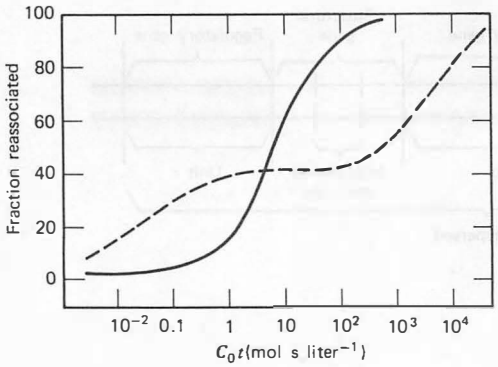


Figure 17.35
Comparison of reassociation kinetics of calf thymus DNA and E. coli DNA.
As opposed to the relative uniformity of DNA from E. coli, the DNA from calf thymus (dotted line) shows the presence of three classes of fragments: those containing unique sequences with $C_0t \sim 10^3$; middle-repetitive sequences with $C_0t \sim 10^{-1}$; and a small highly repetitive class with $C_0t < 10^{-3}$. The presence of repeated sequences makes it possible to renature parts of calf thymus DNA more rapidly than E. coli DNA (solid line) in spite of the much larger size of the genome of the former. Values on the abscissa are obtained by multiplying the total concentration of nucleotides by the renaturation time.

Redrawn from R. J. Britten and D. E. Kohne, *Science*, 161:529, 1968. Copyright 1968 by the American Association for the Advancement of Science.

then allowing them to reassociate and form double-stranded molecules (Figure 17.17). The kinetics of reassociation obey a single second-order equation, indicating that all the sequences in the procaryotic genomes occurs as single copy. When a eucaryotic DNA (mouse DNA) was first studied by this method, unexpected results were obtained, which lead to the realization that eucaryotic DNAs contain reiterated sequences. A priori, it was assumed that since mammalian genes are about three orders of magnitude larger than *E. coli* genes, the rates of reassociation of denatured mammalian DNA would be exceedingly slow. Instead it turned out that a fraction of the mouse DNA, the highly repetitive fraction (mouse satellite DNA), reassociated far more rapidly than even the DNAs of small viruses (Figure 17.17). This is reasonable, since the probability that a fragment will encounter a complementary fragment leading to reassociation is proportional to the number of similar sequences repeated in the original DNA prior to fragmentation. The more repetitive the sequence, the more rapid the reassociation. Consequently, the reassociation kinetics of eucaryotic DNAs provided the first evidence for three classes of sequences. The highly repetitive sequences reassociate extremely rapidly. The nonrepetitive reassociate slowly, and the moderately repetitive renature at intermediate rates (Figure 17.35).

17.5 FORMATION OF THE PHOSPHODIESTER BOND IN VIVO

Two distinct but related phenomena are examined in the sections that follow. These are the processes of enzymatic repair of certain randomly introduced changes in the chemical structure of the DNA bases and the process of DNA replication.

DNA repair and particularly DNA replication are very complex processes. After almost a quarter-century of intensive research much is known, but much still remains unclear about their mechanisms. Although key similarities in the mechanisms of DNA replication and repair are discernible among different organisms, a considerable amount of diversity exists in terms of individual detail. This diversity further complicates any attempt to present a simplified and universally applicable model for each of these two processes. To resolve this difficulty the basic mechanistic elements of the substeps of each process are first described and subsequently integrated, using as

an example the *E. coli* replication system. Certain differences between *E. coli* and eucaryotic systems are also pointed out.

DNA-Dependent DNA Polymerase

The common denominator between the processes of DNA replication and repair is the enzymatically catalyzed synthesis of DNA polynucleotide segments, which can be assembled with preexisting polynucleotides, leading to products of repair or replication. The synthesis of these polynucleotide segments is catalyzed by the enzyme DNA-dependent DNA polymerase, which in the case of *E. coli* has been isolated in three distinct forms, the polymerases I, II, and III listed in Table 17.6. The DNA polymerases are characterized by a 3′ ⟶ 5′-exonuclease activity in addition to synthetic activities. Polymerases I and III are 5′ ⟶ 3′-exonucleases. The involvement of all these enzymatic activities in the processes of repair and replication will be apparent shortly.

The synthetic activity of DNA polymerase can be described by referring to Figure 17.36, in which two complementary DNA strands of unequal length are shown. This configuration, in which the shorter strand has a free 3′ terminus, is essential for the function of DNA polymerase. The enzyme catalyzes the addition of free 5′-deoxynucleotide triphosphates to the 3′ terminus of the short strand, the *primer*. The term *primer* applies to the initial terminus of a molecule, in this instance the 3′-polynucleotide end, onto which additional monomeric units can be added stepwise to yield the final product. The free portion of the longer complementary strand is used as a *template* to direct the condensation of selected 5′-deoxynucleotides onto the growing primer. In the present context the term *template* refers to a single strand of nucleic acid, which

Table 17.6 DNA Polymerase I, II, and III of *E. coli*

	Polymerase		
Properties	*I*	*II*	*III*
Molecular weight	110,000	120,000	180,000
Molecules per cell	400	100	10
Polymerization activity (turnover number)	1,000	50	15,000
Exonuclease activity 3′ ⟶ 5′	Active	Active	Active
Exonuclease activity 5′ ⟶ 3′	Active	Inactive	Active

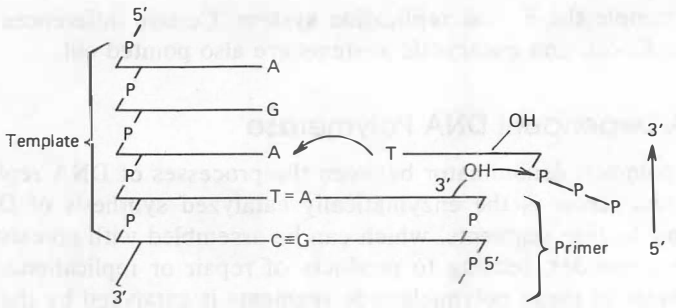


Figure 17.36
Synthetic activity of DNA polymerase.
DNA polymerase catalyzes the polymerization of nucleotides in the 5′ ⟶ 3′ direction. A phosphodiester bond is formed between a free 3′-hydroxyl group of the strand undergoing elongation (the primer) and an incoming deoxynucleoside 5′-triphosphate. Pyrophosphate is eliminated.
Redrawn from A. Kornberg, Science 163:1410, 1969. Copyright 1969 by the American Association for the Advancement of Science.

provides the specific information necessary for the synthesis of a complementary strand. DNA polymerase requires both a primer and a template in order to function. The primer provides a site for the polymerization to begin, and the template provides the information that determines the precise nucleotide sequence of the new polymer.

The DNA polymerase-catalyzed reaction permits the selection of 5′-deoxynucleoside triphosphates, one at a time, with a base complementary to that present in the corresponding position of the template. The specificity of the polymerase reaction with respect to the template is vested in the strong association of each of the bases of the template with their normal complementary partners present in the cell as free 5′-deoxyribonucleotides. Strong binding between complementary bases is apparently achieved because the bases become confined within custom-fitted cages created by appropriate hydrophobic regions of the DNA polymerase. As a result the reading of the template is extremely accurate. In addition, the fidelity of the reading is probably enhanced because the 3′ ⟶ 5′ exonuclease activity of the polymerase may be used for proofreading the bases, selected by the enzyme, for possible errors. Specifically, if a 5′-deoxyribonucleotide which is not complementary with the corresponding base on the template is erroneously condensed with the primer, the enzyme can temporarily reverse its synthetic activity and hydrolyze the phosphodiester bond formed between the primer and the erroneous base. Thus, in effect the enzyme can retrace the path it covers and remove any erroneously introduced mismatched bases. Because of these precautions, rates of error are extremely low, generally in the order of a few mispairings per billion of added pairs.

The polymerase has well-defined selectivities also in a different sense. Only the 3′ terminus of a strand can be used for priming. Therefore the enzyme can elongate a strand only in the 5′ ⟶ 3′ direction, as indicated in Figure 17.36. The 5′ terminus of the strand is rejected as a primer because the polymerase is unable to elongate a polynucleotide in the opposite 3′ ⟶ 5′ direction.

17.6 MUTATION AND REPAIR OF DNA

Mutations

One of the fundamental requirements for a structure that serves as a permanent depository of genetic information is extreme stability. Such stability is essential, at least in terms of those characteristics of the structure that code for the genetic information. Therefore a prerequisite for the structure of DNA is extreme stability in its base content and in its sequence, in which hereditary information is encoded. Yet the structure of the DNA bases is not totally exempt from gradual change. Normally, changes occur infrequently and then affect very few bases, but nevertheless they do take place. Chemical or irradiation-induced reactions may modify the structure of some bases or may disrupt phosphodiester bonds and sever the strands. Errors may also occur during the processes of replication and strand recombination, leading to the incorporation of one or more erroneous bases into a new strand. In almost every instance, however, a few cycles of DNA replication are required before a modification in the structure of a base can lead to irreversible damage, that is, DNA polymerase must use the polynucleotide initially damaged as a template for the synthesis of a complementary strand for the initial change to become permanent. As Figure 17.37 suggests, use of the damaged strand as template extends the damage from a change of a single base to a change of a complete base pair and subsequent replication perpetuates the change.

Since the properties of cells and of the organisms constructed from them ultimately depend on the DNA sequences of their genes,

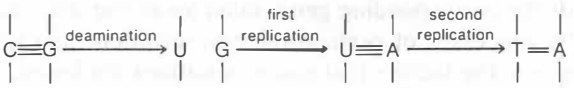


Figure 17.37
Mutation perpetuated by replication.
Mutations introduced on a DNA strand, such as the replacement of a cytosine residue by a uracil residue resulting from deamination of cytosine, extend to both strands when the damaged strand is used as template during replication. In the first round of replication uracil selects adenine as the complementary base. In the second round of replication uracil is replaced by thymine. Similar events occur when the other bases are altered.

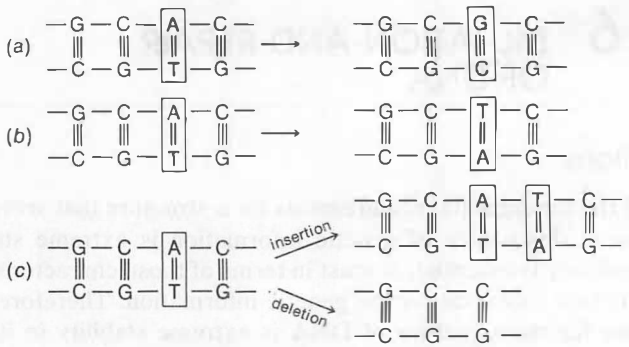


Figure 17.38
Mutations.
Mutations can be classified as transitions, transversions, and frame shift. Bases undergoing mutation are shown in boxes. (a) Transitions: A purine–pyrimidine base pair is replaced by another. This mutation occurs spontaneously, possibly as a result of adenine enolization or can be induced chemically by such compounds as 5-bromouracil or nitrous acid. (b) Transversions: A purine–pyrimidine base pair is replaced by a pyrimidine–purine pair. This mutation occurs spontaneously and is common in man. About one-half the mutations in hemoglobin are of this type. (c) Frame shift: This mutation results from insertion or deletion of a base pair. Insertions can be caused by mutagens such as acridines, proflavin and ethidium bromide. Deletions are caused by deaminating agents. Alteration of bases by these agents prevents pairing.

irreversible alterations in a few DNA base pairs can cause substantial changes in the corresponding organism. These changes, referred to as mutations, may be hidden or visible, that is, phenotypically expressed. Therefore, a *mutation* may be defined as a stable change in the DNA structure of a gene, which is frequently expressed as a phenotypic change in the corresponding organism. Mutations may be classified, depending on their origin, into two categories: base substitutions and frame shift mutations. Base substitutions include *transitions,* substitutions of one purine–pyrimidine pair by another, and *transversions,* substitutions of a purine–pyrimidine pair by a pyrimidine–purine pair. *Frame shift* mutations, which are the most radical, are the result of either the insertion of a new base pair or the deletion of a base pair or a block of base pairs from the DNA base sequence of the gene. These changes are illustrated in Figure 17.38.

Mutagens

A more systematic coverage of the subject of mutations, especially with respect to the expression of a mutation as a change in the product of the corresponding gene, must await the detailed description of the processes of replication, transcription, and translation. In this section, the factors that cause mutations are listed, and a few examples of structural DNA changes brought about by these factors are given. Irradiation and certain chemical compounds are recognized as the main mutagens. Some rare events of incorporation of erroneous bases by DNA polymerase also lead to mutations.

Chemical Modification of the Bases

The bases in DNA are sensitive to the action of numerous chemicals. Among them are nitrous acid (HNO_2), hydroxylamine (NH_2OH), and various alkylating agents such as dimethyl sulfate and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. Chemical modifications of

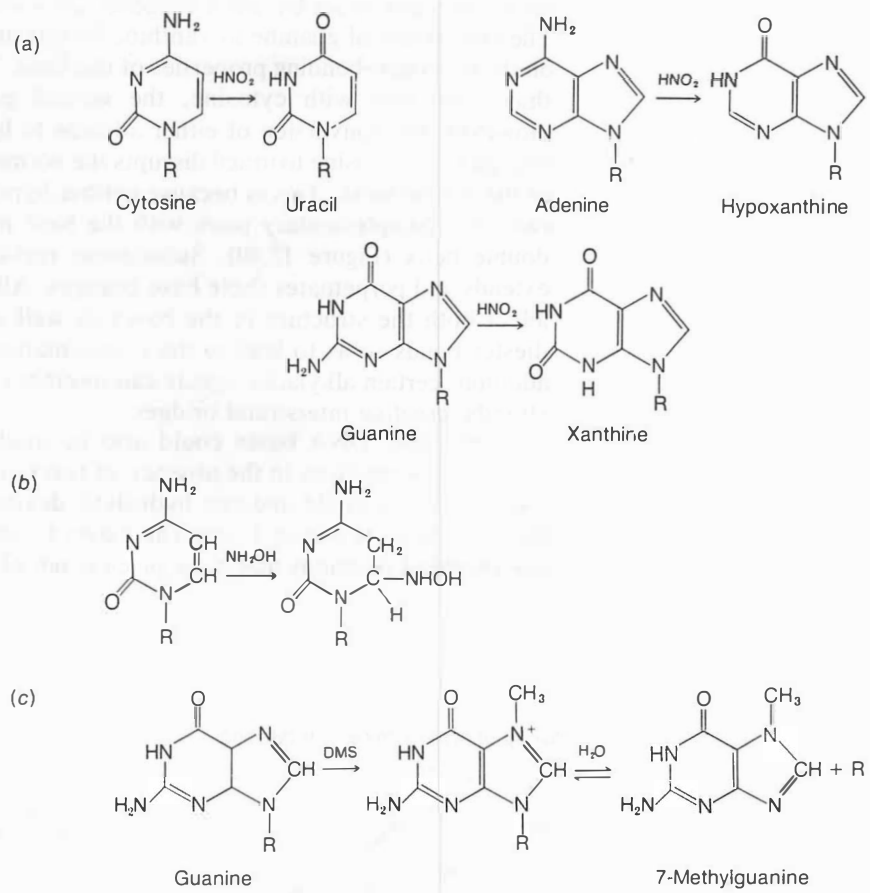
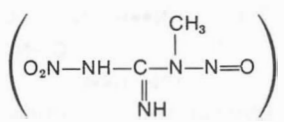


Figure 17.39
Reactions of various mutagens.
(a) Deamination by nitrous acid (HNO_2) converts cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine. (b) Reaction of bases with hydroxylamine (NH_2OH) as illustrated by the action of this reagent on cytosine. (c) Alkylations of guanine by dimethyl sulfate (DMS). The formation of a quaternary nitrogen destabilizes the deoxyriboside bond and releases deoxyribose. Among the effective agents for methylation of the bases are certain nitrosoguanidines such as N-methyl-N'-nitro-N-nitrosoguanidine.



bases, brought about by these reagents are shown in Figure 17.39. The conversion of guanine to xanthine by nitrous acid has no effect on the hydrogen-bonding properties of this base. The new base, xanthine, can pair with cytosine, the normal partner of guanine. However, the conversion of either adenine to hypoxanthine or the change from cytosine to uracil disrupts the normal hydrogen bonding of the double helix. This is because neither hypoxanthine nor uracil can form complementary pairs with the base present in the initial double helix (Figure 17.40). Subsequent replication of the DNA extends and perpetuates these base changes. Alkylating agents may affect both the structure of the bases as well as disrupt phosphodiester bonds so as to lead to the fragmentation of the strands. In addition, certain alkylating agents can interact covalently with both strands, creating interstrand bridges.

In principle, DNA bases could also be modified as a result of spontaneous reactions in the absence of reactive chemicals. For instance, adenine could undergo hydrolytic deamination to hypoxanthine as shown in Figure 17.41. The extent to which such spontaneous chemical reactions may take place is not established.

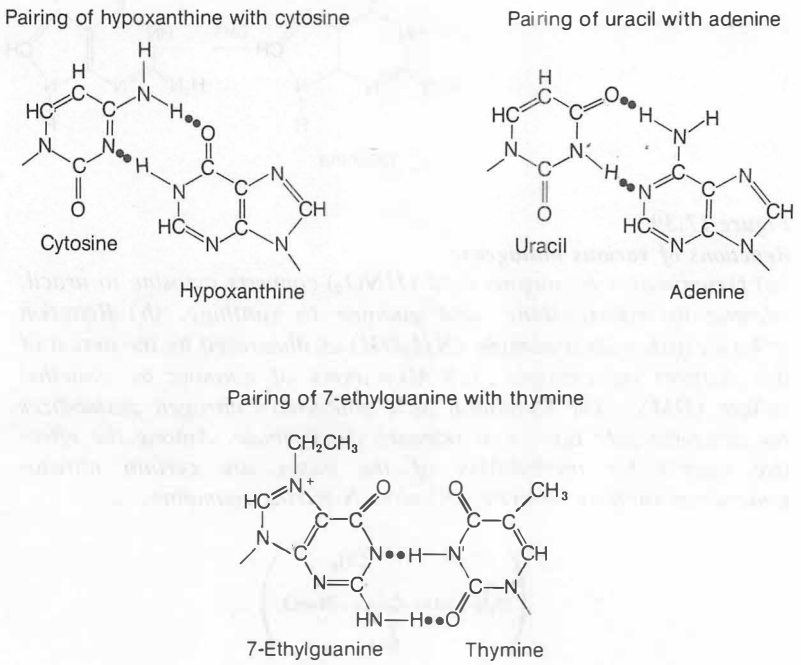


Figure 17.40
Chemical modifications that alter the hydrogen-bonding properties of the bases.
Hypoxanthine, obtained by deamination of adenine, has different hydrogen-bonding properties from adenine, for example, it pairs with cytosine. Similarly, uracil obtained from cytosine, has a different hydrogen-bonding specificity than cytosine and pairs with adenine. Alkylation of guanine modifies the hydrogen-bonding properties of the base.

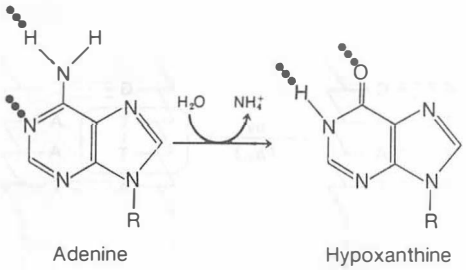


Figure 17.41
Hydrolytic deamination of adenine.

Radiation Damage

Both ultraviolet as well as x-ray irradiation are generally very effective means of producing mutations. The bases normally exist as keto or amino forms in equilibrium, with only minor amounts of the enol or the imino structures, as shown in Figure 17.42. Radiation energy absorbed by the bases tends to shift the equilibrium to the minor forms. The minor forms, however, cannot pair with the normal partners of the bases. For example, the imino form of adenine pairs with cytosine instead of thymine, the normal adenine partner. This atypical base pairing is shown in Figure 17.43. It has been suggested that the existence of increased amounts of the enol forms of the bases at the moment of replication increases the frequency of mutations of the newly synthesized DNA strand because the enol forms

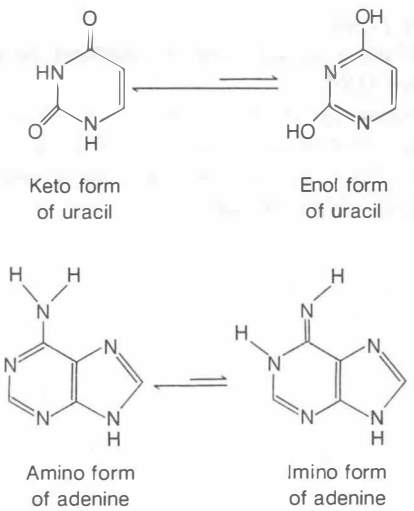


Figure 17.42
Tautomeric equilibrium of the bases.
An equilibrium is present between the tautomeric forms of purines. The same is true for pyrimidines. The interconversion between the two tautomeric forms is a process involving the interchange of protons between different positions of the molecule. At physiological pH the amino and the keto forms are strongly favored by the equilibrium.

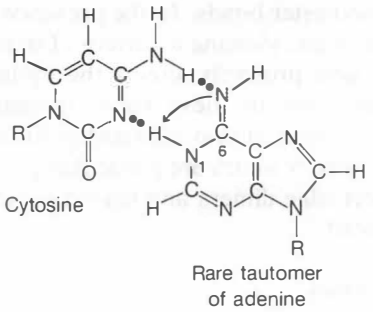
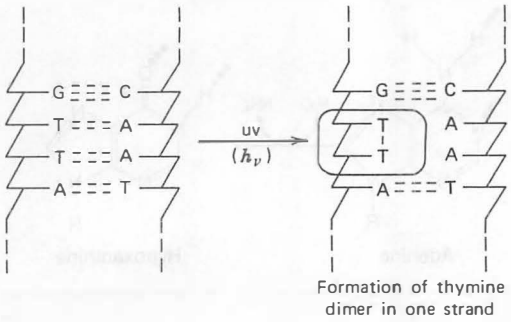


Figure 17.43
Base complementarity properties of the minor tautomeric form of adenine.
The hydrogen-bonding properties of adenine are fundamentally changed when adenine acquires the minor tautomeric form. The pairing with cytosine, shown in this figure, is very atypical of the normal properties of this base.



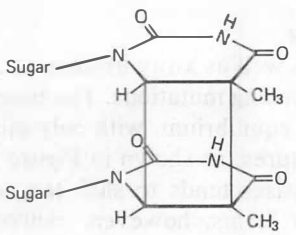


Figure 17.44
Dimerization of adjacent pyrimidines in irradiated DNA.
A residue of thymine which is activated by the absorption of ultraviolet light can react with a second neighboring thymine and form a thymine dimer.

select new bases that pair with them rather than the normal hydrogen-bonding partners of the more predominant keto forms.

Exposure of DNA to high energy radiation (x-rays or γ-rays) may also bring about direct modifications in the structure of the bases. Intermediates produced by electron expulsion can be rearranged, leading to the opening of the heterocyclic rings of the bases and the disruption of phosphodiester bonds. In the presence of oxygen additional reactions take place, yielding a variety of oxidation products.

Irradiation by uv light primarily affects the pyrimidines. Activation of the ethylene bond of these bases frequently leads to a photochemical dimerization of two adjacent pyrimidines, as shown in Figure 17.44. Thymine residues are particularly susceptible to this reaction, although cytosine dimers and thymine–cytosine combinations are also produced.

DNA Polymerase Errors

When the appropriate deoxyribonucleotides are available, DNA-dependent DNA polymerase generally functions with a very high degree of fidelity. Not all DNA polymerases, however, are equally discriminating. For instance, the RNA-dependent DNA polymerase (a polymerase that uses RNA as a template) associated with a virus causing a form of leukemia in birds can make numerous errors in selecting nucleotides complementary to the bases of the template. Perhaps more than one erroneous base per 1,000 correctly chosen

nucleotides might be incorporated into DNA by this low specificity polymerase. Furthermore, even the most discriminating DNA polymerases are unable to distinguish between the normal deoxyribonucleotide triphosphate substrates and other nucleotides with very similar structures.

On the basis of the tendency of DNA polymerase to accept, in place of the normal substrates, structural analogs of the common bases, certain mutations can be introduced into DNA by design. For instance, 2-aminopurine, incorporated instead of adenine into a newly synthesized DNA strand, can associate with cytosine and produce an A–T ⟶ G–C transition. A somewhat more complex example is provided by 5-bromouracil. This base, in the form of the corresponding deoxynucleoside triphosphate, can be incorporated into a strand in place of thymine. However, the equilibrium between the enol and the keto forms for these two bases allows for the formation of a somewhat higher proportion of the enol form in 5-bromouracil than in thymine. This occurs presumably because of the higher electronegative nature of the bromine atom in comparison to the corresponding methyl group in thymine. Because the enol form of 5-bromouracil pairs with guanine, as shown in Figure 17.45, the substitution of thymine by bromouracil produces an A–T ⟶ G–C transition.

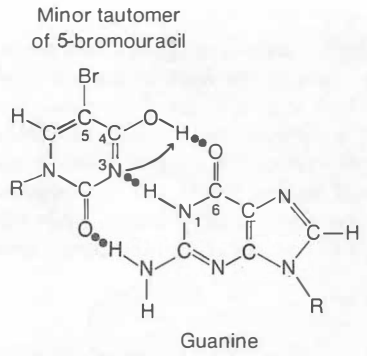


Figure 17.45
Hydrogen-bonding properties of the minor enol form of 5-bromouracil.
The enol form of 5-bromouracil, an analog of thymine, pairs with guanine instead of adenine, the normal partner of thymine.

The Stretching of the Double Helix

Certain organic compounds, which are characterized by planar aromatic ring structures of appropriate size, can be inserted between base pairs in double-stranded DNA. This process is referred to as *intercalation*. During intercalation neighboring base pairs in DNA are separated to allow for the insertion of the intercalating ring system, causing an elongation of the double helix by stretching. The continuity of the base sequences in DNA is disrupted, and the reading of the bases by the DNA polymerase produces a new strand, with an additional base inserted near the site of intercalation. The resulting mutation is referred to as a *frame shift*. *Acridines, ethidium bromide,* and other intercalators are known to be effective *frame-shift mutagens* (Figure 17.46).

Clin. Corr. 17.1 discusses mutations and the etiology of cancer.

Repair of DNA

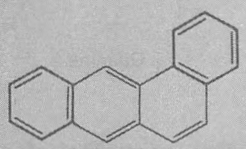
Mutations may be necessary for effective evolutionary response to environmental change. Procaryotes, especially those that use eucaryotes as hosts, are able to generate new forms rapidly. This flexibility is essential for survival because eucaryotic hosts become

CLIN. CORR. 17.1 MUTATIONS AND THE ETIOLOGY OF CANCER

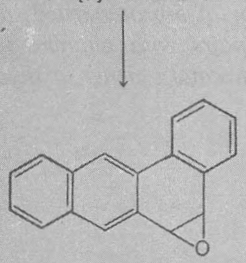
Considerable progress in our understanding of the etiology of cancer has been made in recent years by our ever-increasing realization that long-term exposure to certain chemicals leads to various forms of cancer. Some experts are now suggesting that the great majority of cancers are in fact triggered by environmental factors.

Carcinogenic (cancer-causing) compounds are not only introduced into the environment by the increasing use of new

chemicals in industrial applications but are also present in the form of natural products. For instance, the *alfatoxins*, produced by certain molds, and *benz*[a]*anthracene*, present in cigarette smoke and charcoal broiled foods, are carcinogenic. Some carcinogens act directly, while others, such as benz[*a*]anthracene, must



Benz[a]anthracene

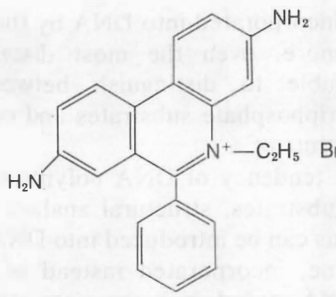


5,6-Epoxide (carcinogenic)

undergo prior hydroxylation by arylhydroxylases, present mainly in the liver, before their carcinogenic potential can be expressed.

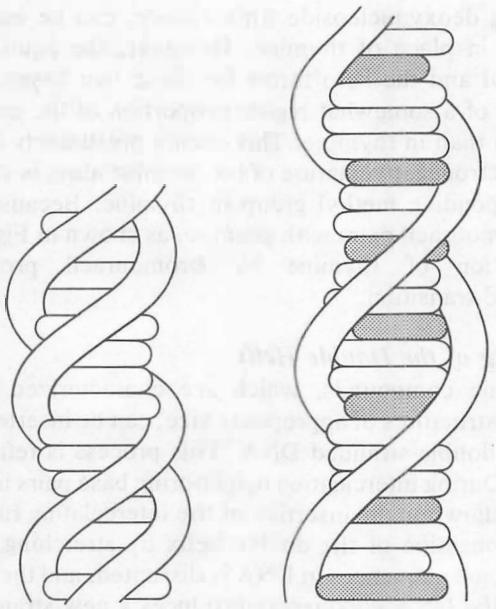
The reactivity of many carcinogenic compounds toward guanine residues results in modification of the guanine structure, usually by alkylation at the N7 position, as well as in breaks of the phosphodiester bond, events that, upon replication, lead to permanent mutations. Chemicals that produce mutations generally turn out to be carcinogenic and vice versa.

The vulnerability of DNA toward alkylating agents, and other chemicals as well, underscores the concerns expressed today by many scientists about the ever-increasing exposure of our environment to new chemicals. What is particularly distressing is that the carcinogenic potential



Ethidium bromide

(*a*)



(*b*)

Figure 17.46

Intercalation between base pairs of the double helix.

The insertion of the planar ring system of intercalators (a) between two adjacent base pairs requires the stretching of the double helix (b). During replication this stretching apparently changes the frame used by DNA polymerase for reading the sequence of nucleotides. Consequently, newly synthesized DNA is frame-shifted. (b-1) The original DNA helix; (b-2) The helix with intercalative binding of ligands.

rapidly immune to the original forms of invading organisms. Adaptability, however, is obtained at high cost, since most random mutations are not beneficial but instead produce deleterious effects.

Because of the generally harmful nature of mutations, mechanisms have also evolved that counteract the effects of these changes. In effect the rate of initial DNA damage is generally much higher than the rate of the expressed lethal or damaging mutations. Obviously the most direct means by which a potential mutation can be reversed is the repair of the affected base prior to replication. Almost all cells have the capacity to detect distortions in the structure of the DNA double helix. Such distortions may result from a permanent interruption of the hydrogen bonding between one or more base pairs which originates from a change in the structure of one of the bases. Two principal mechanisms that lead to the restoration of the normal structure of DNA are presented here.

Excision Repair

One of the better understood repair processes is the repair of ultraviolet-damaged DNA in *E. coli*. This is a multistep process carried out with the coordinated participation of several enzymes, including nucleases, DNA polymerase, and DNA ligase. The process is illustrated in Figure 17.47. The repair is initiated by the recognition of the distortion of the DNA structure by an endonuclease. The enzyme binds to the distorted area and catalyzes the hydrolytic cleavage of a phosphodiester bond in the strand that contains the distortion. A nick is introduced on the 5′ side of the modified base. Once the 5′ end is liberated, a 5′ $\longrightarrow$ 3′ exonuclease begins to remove gradually either a few (*short-patch repair*) or numerous (*long-patch repair*) nucleotides from the 5′ end of the nicked strand.

An example of long-patch repair is the removal of pyrimidine dimers. The dimers are removed by the exonuclease along with a large number of neighboring nucleotides. The resulting gap is then filled by DNA polymerase I, so as to restore the original structure. Short-patch repair takes place under different circumstances, for instance, when a strand is broken by ionizing radiation. As few as two or three nucleotides are removed by the exonuclease prior to the initiation of the DNA polymerase-catalyzed synthesis of the new strand.

The mechanism by which the DNA polymerase I fills the gap in the DNA strands undergoing repair deserves some special attention because in its general features it is applicable to the processes of both DNA repair as well as DNA synthesis. This polymerase exhibits a combination of synthetic and exonucleolytic activities, which are vested into separate active sites of the enzyme. The synthetic site uses as a primer the 3′ terminus of the strand from which the

of new chemicals released into the environment cannot be predicted with confidence even when they appear chemically innocuous toward DNA.

Until recently, tests for carcinogenicity, that is, the ability of a substance to cause cancer, required the use of large numbers of experimental animals to which high doses of the suspected carcinogen were administered over a long period of time. Such tests, which are time-consuming as well as expensive, are the only approach still available for testing carcinogenicity directly. Recently, however, a much simpler and much more inexpensive indirect test for carcinogenicity was developed. This test is based on the premise that carcinogenicity and mutagenicity are essentially manifestations of the same underlying phenomenon, the structural modification of DNA. The test measures the rate of mutation that bacteria undergo when exposed to chemicals suspected to be carcinogens.

A major criticism advanced against the test is that the assumption of an equivalence between mutagenicity and carcinogenicity is not always valid. Because of the unusually large economic implication of labeling a chemical with widespread use as a potential carcinogen, extreme scrutiny regarding the reliability of any proposed diagnostic test would not be surprising. Yet, certain exceptions notwithstanding, the great majority of chemicals tested has reinforced the view that a good correlation exists between the tendency of a chemical to produce bacterial mutations and animal cancer. Furthermore, even the direct and very costly tests for carcinogenicity, in which large numbers of animals are used, have not completely escaped criticism. The reliability of the test has been questioned because of the relatively large doses of chemicals employed in these tests; doses essential for shortening the long-term chemical exposure of the animals to a practically mangeable period of time.

In addition, the necessity of projecting data from animals, usually rodents, to humans has often been used as an argument against the validity of the test.

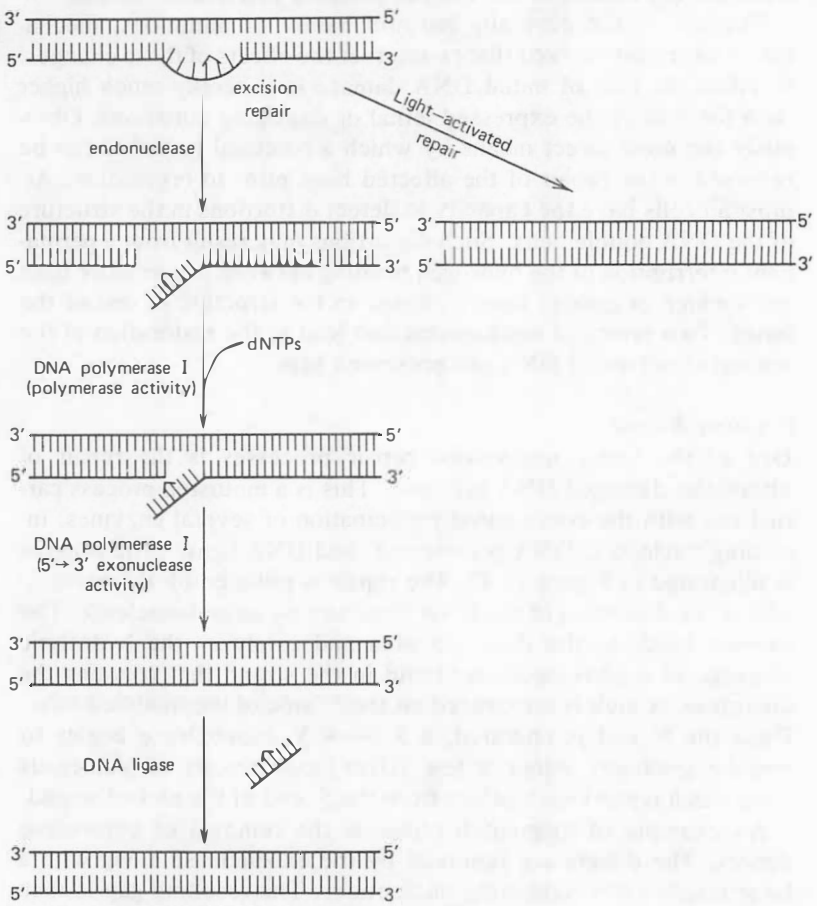


Figure 17.47
Repair of DNA damaged by ultraviolet light.
Excision repair requires the coordinated action of many enzymes. The repair begins with an incision by a specific endonuclease at a point several bases away from the distorted region of DNA that contains the dimer. DNA polymerase I elongates the 3′-OH terminus, using the other DNA strand as a template. Polymerase I also removes the DNA fragments containing the dimer. The final gap is sealed by the action of DNA ligase, and the damaged DNA region is restored to an intact state. Alternatively, the dimerization can be directly reversed by a process catalyzed by light, known as light-activated repair.

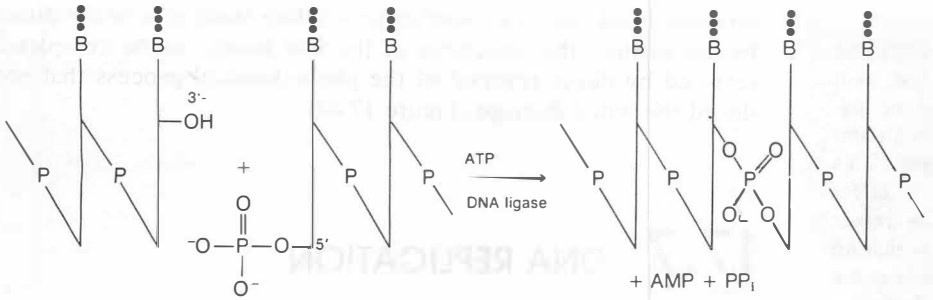


Figure 17.48
The action of DNA ligase.
The enzyme catalyzes the joining of polynucleotide strands that are part of a double-stranded DNA. A single phosphodiester bond is formed between the 3'-hydroxyl and the 5'-phosphate ends of the two strands.

In E. coli cells the energy for the formation of the bond is derived from the cleavage of the pyrophosphate bond of NAD^+. In eucaryotic cells and bacteriophage-infected cells energy is provided by the hydrolysis of the α,β-pyrophosphate bond of ATP.

pyrimidine dimer has been removed. The intact complementary strand serves as the template that directs the incorporation of the nucleotides needed to restore the original structure of the strand undergoing repair. The $3' \longrightarrow 5'$ exonucleolytic activity of polymerase I insures the fidelity of the synthesis as usual. At the same time the distinct $5' \longrightarrow 3'$ hydrolytic site of the enzyme removes nucleotides from the 5′ terminus of the strand undergoing repair. The concerted actions of polymerase I therefore close the gap created by the excision of the pyrimidine dimer until the gap is reduced to the size of a nick, that is, until a single phosphodiester bond remains open. As a result of the combined synthetic–nucleolytic action of polymerase I, the nick can move along the strand undergoing repair (nick translation) until it is finally bridged by the action of DNA ligase, as illustrated in Figure 17.48. Clin. Corr. 17.2 discusses defects in DNA repair which are associated with human disease.

Light-Activated Repair

Photochemical damage in DNA, such as the formation of pyrimidine dimers, can also be reversed by the action of ultraviolet light. Some cells contain a protein that can associate with pyrimidine dimers. Absorption of light activates this protein, which in turn breaks the

CLIN. COR. 17.2 DISEASES OF DNA-REPAIR AND SUSCEPTIBILITY TO CANCER

Defects in DNA repair may lead to a number of diseases including *Xeroderma pigmentosum, Ataxia telangiectasia* and *Fanconi's anemia*. *Xeroderma pigmentosum* is the best understood of these conditions. Those suffering from this affliction are particularly sensitive to the effects of sunlight. Exposure to the sun creates severe skin reactions which range initially from excessive flecking and skin ulceration to the eventual development of skin cancers. The disease is also accompanied by various neurological abnormalities.

Normal mammalian cells can carry out excision repair of DNA. In contrast, skin cells from *Xeroderma pigmentosum* patients are unable to repair the DNA damage produced by ultraviolet light. This defect originates from a reduced effectiveness of the first step of the repair mechanism, that is, the step at which an endonuclease nicks the DNA undergoing repair near a pyrimidine dimer. It has been suggested, based on indirect evidence, that in *Xeroderma pigmentosum* patients this enzyme is defective. However, there are at least five and possibly seven genetic forms of the disease as determined by cell hybridization studies. Therefore, if this suggestion were correct, the endonuclease should contain several components and for each variant the defect should reside in different components.

It is not clear though whether the enzyme itself is defective or whether the excision repair defect is the result of impaired access of the endonuclease to chromosomal DNA. If this is the case, then some yet unidentified control factors needed to expose DNA, in order that they become susceptible to the action of endonuclease, could be missing.

One particularly intriguing aspect of diseases related to defective DNA repair is their possible relationship to carcinogenesis. Although these diseases are rare autosomal recessive conditions, the carriers of the defective genes are relatively common; carriers of *Xeroderma pigmentosum* gene account for as much as 1% of the general population. An interesting recent finding is that not only are those who are afflicted by *Xeroderma pigmentosum* susceptible to cancer but the carriers of the gene have a higher incidence of skin cancer. Also, carriers of the *Ataxia telangiectasia* gene have an increased risk of dying from cancer. These and other similar findings suggest the presence of some subtle defect in DNA repair among

covalent bonds holding together pyrimidine molecules in the dimer. In this manner the structures of the free bases can be completely restored by direct reversal of the photochemical process that produced the initial damage (Figure 17.47).

17.7 DNA REPLICATION

Basic Elements of the Mechanism of Replication

From the very same moment the double-stranded structure of DNA was proposed, it was apparent that this structure could serve as the basis of a mechanism for DNA replication. The complementary structure of the strands was immediately perceived as a characteristic, which, in principle, permitted each one of the strands to serve as a template for the synthesis of a new strand identical to the other strand, as suggested in Figure 17.49. A number of pathways could be easiiy visualized that lead to the synthesis of two new double-stranded helices, identical to one another and to the maternal double helix.

In the more than quarter-century since the double helix was proposed, the correctness of this overall scheme of replication has been solidly established. Our expanding knowledge of the character of DNA replication has also revealed that the simplicity of the basic scheme conceals, in fact, a rather complex set of more intricate

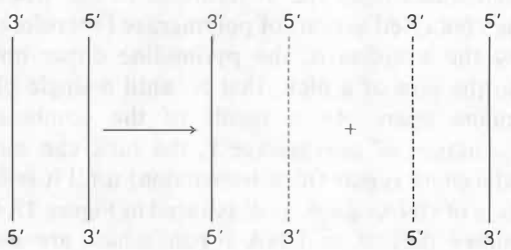


Figure 17.49
Each DNA strand serves as template for the synthesis of a new complementary strand.
Replication of DNA proceeds by a mechanism in which a new DNA strand (indicated by a dashed line) is synthesized that matches each original strand (shown by solid lines).

substeps. The fact is now well appreciated that a multiplicity of enzymes and protein factors participate in the process of replication. It is also clear that before the synthesis of a DNA molecule can be brought to successful completion, the enzymes involved in replication must deal with a variety of topological problems.

These problems originate partly because the DNA-dependent DNA polymerase can synthesize new strands by operating *along the 5′ ⟶ 3′ direction only*, and therefore it is unable to elongate the two *antiparallel* strands of the helix in the *same macroscopic direction*. Also, the replication of DNA cannot proceed unless the complementary strands are separated at an early stage of the synthesis. Separation requires the commitment of energy for disrupting the thermodynamically favorable double-helical arrangement and the unwinding of a highly twisted double helix at extremely rapid rates. As if these difficulties were not enough, double-stranded DNA is normally a topologically closed domain which, unless properly modified, will not tolerate strand unwinding to any appreciable degree. Obviously, these multiple difficulties are enzymatically resolved before the replication of DNA can take place.

In order to ease the complexity of describing the process of DNA replication, the substeps of this process will be described first.

the carriers of these genes. Further research on the predisposition of these carriers to cancer might provide some clues regarding the mechanism of carcinogenesis for some types of cancer.

Replication Is Semiconservative

The concept that DNA strands are separable and that new strands, complementary to the preexisting strands, can be assembled from free nucleotides on each separate strand, is not new. In a macroscopic sense three possibilities by which information transfer could take place during replication were initially visualized as indicated in Figure 17.50. *Conservative* replication could in principle yield a product consisting of a double helix of the original two strands and a daughter DNA consisting of completely newly synthesized chains. A second possibility, labeled *dispersive*, would have resulted if the nucleotides of the parental DNA were randomly scattered along the strands of the newly synthesized DNA.

The synthesis of DNA eventually proved to be a *semiconservative* process. After each round of replication, the structure of parental DNA is found to preserve one of its own original strands combined with a newly synthesized complementary polynucleotide.

The semiconservative nature of replication was elegantly suggested by a classic experiment which allowed the physical separation and identification of the parental and the newly synthesized strands. For this experiment *E. coli* was grown in a medium containing [^{15}N]ammonium chloride as the exclusive source of nitrogen. Several cell divisions were allowed to occur during which the naturally occurring ^{14}N in the DNA of *E. coli* was, for all practical pur-

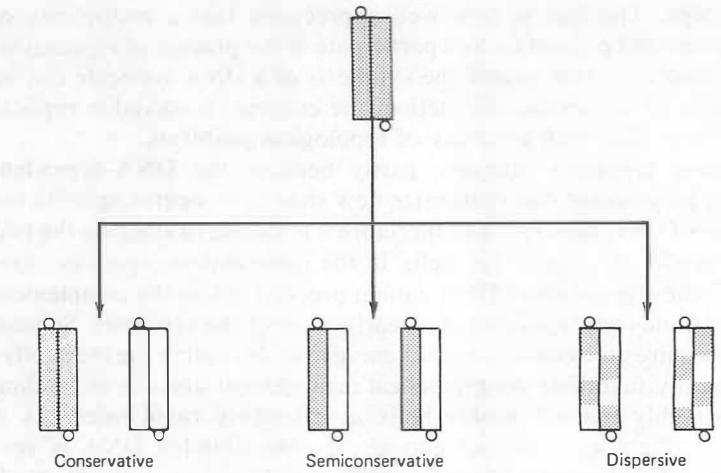


Figure 17.50
Three possible types of DNA replication.
Replication has been shown to occur exclusively according to the semiconservative model, that is, after each round of replication one of the parental strands is maintained intact, and it combines with one newly synthesized complementary strand.

poses, replaced by the heavier ^{15}N isotope. The ^{14}N-containing nutrient was then added, and cells were removed at appropriate intervals. The DNA of these cells was extracted, and the ratios of ^{14}N ^{15}N content were determined by equilibrium density gradient centrifugation. The separation between ^{14}N and ^{15}N DNA was achieved based on the lower density of DNA, which contained the lighter isotope. In subsequent experiments, the newly synthesized DNA was thermally denatured and the individual strands were completely separated. The results, shown in Figure 17.51, demonstrated that daughter DNA molecules consisted of two strands with different densities, corresponding to the densities of single-stranded polynucleotides containing exclusively ^{14}N or ^{15}N.

Clearly, the synthetic activity of DNA polymerase makes it possible for the enzyme to synthesize new complementary DNA strands by using in turn each parental DNA strand as template.

A Primer Is Required

The semiconservative nature of DNA replication requires that each strand serve as a DNA polymerase template for the synthesis of a new complementary strand. The DNA polymerase, however, in spite of its effectiveness as a polynucleotide-elongating enzyme, cannot initiate the synthesis of a strand because it is unable to assemble the first few critical nucleotides. For this reason this enzyme needs a primer that varies in size between perhaps a dozen nucleotides in eucaryotes to approximately five times as many nucleotides in bacteria. This primer is synthesized by other enzymes, as indicated in Figure 17.52.

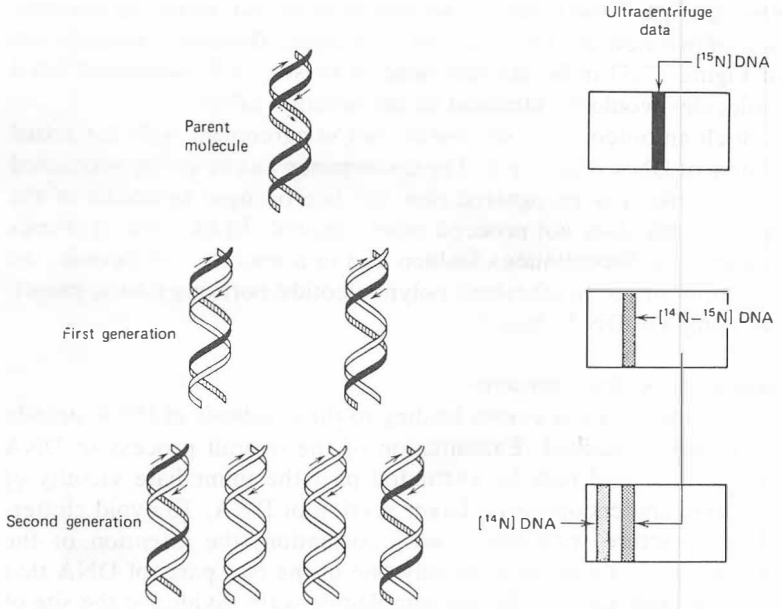


Figure 17.51
Semiconservative replication of DNA.
Schematic representation of the experiment of Meselson and Stahl that demonstrated semiconservative replication of DNA. This model of replication requires that, if the parent molecule (shown in black) contains ^{15}N, each of the molecules produced during the first generation contain ^{15}N in one strand and ^{14}N in the other. Furthermore, in the second generation two molecules must contain only ^{14}N, and two molecules must contain equal amounts of ^{14}N and ^{15}N. The results of separating DNA molecules from successive generations, shown on the right, are consistent with this model.

In some systems, such as phage M13, the priming enzyme has activity characteristic of an *RNA polymerase*, that is, the nucleotides selected and condensed on the template to form the primer are ribonucleotides rather than deoxyribonucleotides. In other systems, including phage ϕX, initiation is achieved by the action of a *primase*. An interesting characteristic of this enzyme, is that it is not particularly discriminating between 5′-ribo- and 5′-deoxyribonucleotides, both of which are selected for the formation of the primer.

Once the primers have been synthesized, by whichever means, the DNA polymerase can move in and take over the process of synthesis. Before the process of replication can be completed, the primers must of course be enzymatically hydrolyzed, leaving behind only the newly synthesized polydeoxyribonucleotide portions.

Both Strands of DNA Serve as Templates Concurrently

In the preceding section the events leading to the initiation of DNA synthesis by DNA polymerase were examined. Attention was directed to one of the two parental DNA strands used as template. In fact, however, initiating events occur at both strands almost concurrently. At first sight this would appear to generate some problems of geometry. Specifically, if a single initiation site is considered, and if the synthesis is assumed to continue until each template is com-

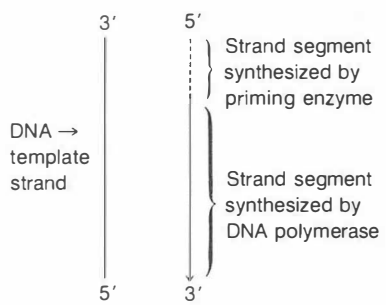


Figure 17.52
Synthesis of primer for DNA replication.
The primer (dashed line) is synthesized by specific enzymes. The existence of a primer permits new DNA (solid line) to be synthesized after which the primer is excised.

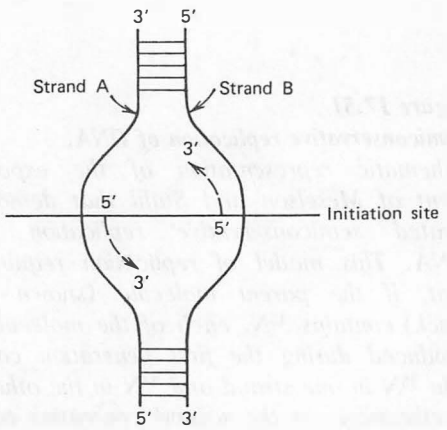


Figure 17.53
Both DNA strands serve as templates for DNA synthesis.
Each DNA must serve as a template for DNA synthesis. The new DNA can be synthesized only in the 5′ ⟶ 3′ direction. For these reasons if only a single initiation site were considered, the result of the synthesis would be the formation of two new nonidentical double-stranded DNA molecules (one above and one below the initiation origin). Also, the upper part of strand A and the lower part of strand B could not have been used as templates. More than a single initiation site is involved.

pletely copied, the result of the synthesis would simply be the creation of *two* new double-stranded molecules. However, examination of Figure 17.53 indicates that none of these two hypothetical DNA molecules would be identical to the parental DNA.

Such an outcome is, of course, not in agreement with the actual course of DNA replication. The discrepancy can be easily accounted for because it is recognized that the microscopic synthesis of the new strands does not proceed uninterrupted. In fact, the synthesis occurs in a discontinuous fashion and in a manner that permits the assembly of the synthesized polynucleotide portions into appropriate complete DNA strands.

Synthesis Is Discontinuous

So far only the early events leading to the synthesis of DNA strands have been described. Examination of the overall process of DNA synthesis should now be expanded past the immediate vicinity of initiation and encompass a larger section of DNA. To avoid cluttering the picture with redundant information, the attention of the reader should be focused on only one of the two parts of DNA that would be generated if the macromolecule were divided at the site of chain initiation, as indicated in Figure 17.54. It may be expected that the synthetic events occurring at this part of the molecule (indicated by solid lines), are of the same detailed nature as those occurring on the other site (dashed lines).

A prerequisite for the semiconservative mechanism of replication is that the two complementary strands of DNA gradually separate as the synthesis of new strands takes place. The mechanics of this separation is addressed later, but it may be apparent that as a result of separating the strands at an interior position, two topologically equivalent forks are created at the point of diversion of the two strands.

The observation to be emphasized presently is that DNA polymerase acts in a discontinuous manner, that is, along each DNA molecule there are numerous points at which primers are formed. How these points are selected is not precisely known, but, at least in *E. coli* and some of its phages, they appear to be determined through the action of *prepriming proteins.* After promoting initiation at one point, these proteins may move along the template strand, assisted by their own ATPase activity. In this manner the prepriming proteins can reach an adjacent initiation site and promote the synthesis of the next polynucleotide segment and so on until a large DNA segment is completed.

The segments built by DNA polymerase upon each primer, which are known as Okazaki fragments, vary in size from approximately 100 deoxyribonucleotides in the case of eucaryotes to about 10 times

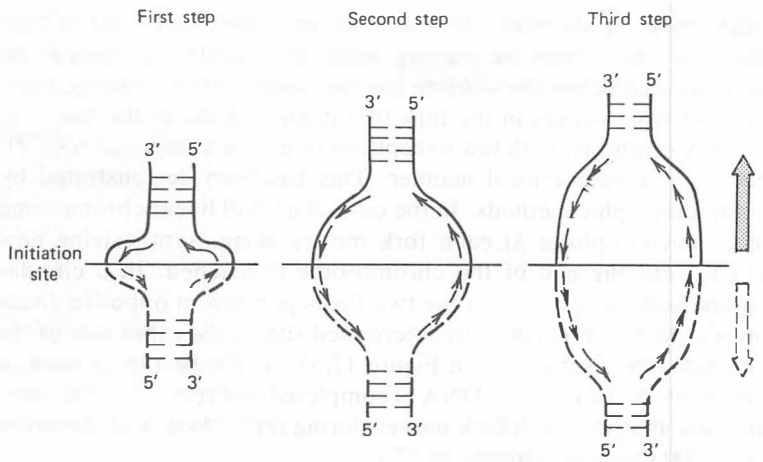


Figure 17.54
Discontinuous synthesis of DNA.
In this figure emphasis is placed on the synthetic events occurring at only one side of the initiation site (solid line). The two complementary strands of DNA separate as the discontinuous synthesis of small DNA segments takes place on both strands. After excision of the primers, the excised parts are repaired, and the segments are joined together. Although the segments are clearly synthesized in opposite directions on the two strands, the overall macroscopic impression is that the DNA grows in the single direction suggested by the solid arrow on the right.

as long in the case of bacteria. Once the small segments of the new DNA strands are synthesized on both strands of a fork (upper part of Figure 17.54), the fork opens up further, and the same process of synthesis can be repeated.

Shortly after synthesis, the primer portions of the Okazaki fragments are excised by the 5′ ⟶ 3′ exonuclease activity of DNA polymerase. The DNA polymerase therefore serves both as an exonuclease as well as a repair enzyme in a manner conceptually similar to that by which photochemically damaged DNA is repaired. The nicks, which remain after repair, are closed by DNA ligase.

This discontinuous mechanism compensates for the inability of DNA polymerase to synthesize strands in the 3′ ⟶ 5′ direction. By synthesizing portions of DNA strands only in the 5′ ⟶ 3′ direction on both antiparallel strands of the parental DNA, the polymerase is able to produce *the illusion that both strands are concurrently elongated in the same macroscopic direction.* In Figure 17.54 this direction is indicated by a large solid arrow. It should be noted that the first strand synthesized, often referred to as the *leading* strand, could alternatively be synthesized *continuously.* It is the second strand, the *lagging* strand, that *must be synthesized discontinuously* to make allowance for the inability of DNA polymerase to elongate primers in the 3′ ⟶ 5′ direction.

Macroscopic Synthesis Is Bidirectional

Examination of Figure 17.54 indicates that at the site of initiation of DNA synthesis two identical forks are created. Therefore two possibilities exist for the synthesis of DNA: the process may occur at only one fork and proceed in a single direction, as shown by the thick

solid arrow, or alternatively it may occur at both forks and in both directions away from the starting point. The events occurring in the forks located below the starting line are simply a mirror image repetition of what occurs in the fork that is located above the line.

DNA synthesis, with few exceptions (e.g., bacteriophage ϕX174), occurs in a bidirectional manner. This has been demonstrated by autoradiographic methods. In the case of a small linear chromosome (e.g., bacteriophage λ) each fork moves along, synthesizing new DNA, until the end of the chromosome is reached. In a circular chromosome (e.g., *E. coli*) the two forks proceed in opposite directions until they meet at a predetermined site on the other side of the chromosome, as depicted in Figure 17.55. As the two forks meet, a new copy of the parental DNA is completed and released. The average rate at which each fork moves during replication is of the order of 50,000 bases per minute at 37°C.

In spite of the efficiency of DNA polymerase as a synthetic enzyme, *mammalian chromosomes* are much too large to be made *in* one piece. Based on the rate at which DNA polymerase can incorporate nucleotides onto a mammalian primer, it may be calculated that a large mammalian DNA molecule would have taken as much as 1 to 2 months to be synthesized. Instead, in most instances replication is completed within a few hours. This is possible because DNA polymerase molecules initiate bidirectional synthesis, not in one, but at several initiation points along the chromosome. The DNA segments between two initiation points are termed *replicons*. Therefore, for a DNA molecule that contains 1,000 replicons, replication may proceed simultaneously at as many as 2,000 forks. At each of these forks, strands are being replicated as Okazaki fragments. The reader should note the distinction between a replicon and an Okazaki frag-

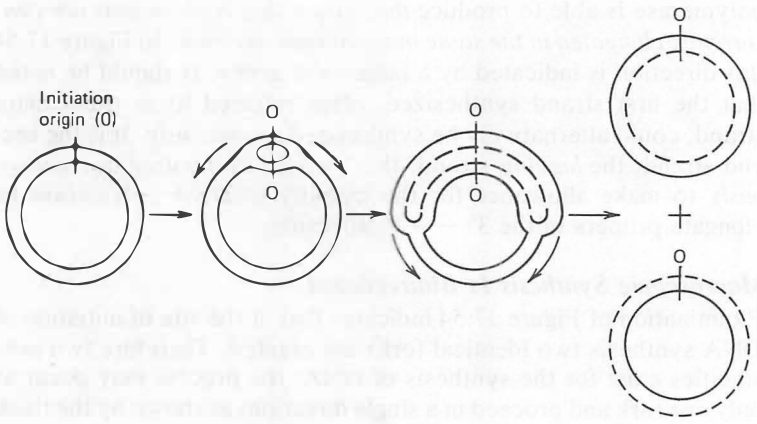


Figure 17.55
Bidirectional replication of a circular chromosome.
Replication starts at a fixed initiation origin (O), and goes in opposite directions until the replication forks meet. Newly synthesized strands are indicated by a dashed line.

ment. Replicons vary considerably in size and may extend across as many as 40,000 nucleotide pairs, whereas an Okazaki fragment is a much shorter DNA intermediate, which in eucaryotes typically consists of about 100 nucleotides. Each mammalian chromosome may use for replication as many as several hundred replicons of different sizes.

Strands Must Unwind and Separate

One aspect of replication, which in the preceding presentation has been taken for granted, is the necessity of separating the strands of the parental DNA prior to the synthesis of new strands. Separation is essential because the bases of each template must be made accessible to the complementary deoxyribonucleotides from which the new strands are constructed. The overall process of separation consists of a number of enzymatically catalyzed, coordinated steps, including the local unwinding of the helix, and the nicking and rejoining of the strands necessary for the continuation of the unwinding process. Once the strands are unwound, they must be kept separate so that they can operate freely as templates.

Helicase. The fact that most DNAs are circular supertwisted molecules facilitates the unwinding of the helix. More precisely, they contain a net deficit of helical turns in comparison to the corresponding "relaxed" molecules. The deficit introduces into the double helix a tendency toward partial unwinding of the two strands. This tendency is even more pronounced in regions richer in A-T pairs, which are intrinsically less stable than G-C pairs. Therefore A-T-rich regions are susceptible to local melting, that is, unwinding and partial separation of the strands of the double helix.

In spite of this tendency, the cell has to resort to the services of specialized proteins to accomplish the rapid orderly unwinding of the strands. These proteins, which in the specialized literature are usually named as products of their corresponding genes, separate DNA strands in advance of the moving replication fork. For the *E. coli* system the name *helicase* has also been used to describe these proteins. The free energy required for the separation of the strands of the helix appear to be provided by the hydrolysis of two molecules of ATP, which are utilized per each base pair melted.

Binding proteins. Once the strands have been separated, the single-stranded regions generated are stabilized by specific proteins, the *binding proteins,* which have been identified in both procaryotic and eucaryotic systems. The hallmark of binding proteins is that they keep the separated DNA strands apart by maintaining the "melting temperature" of the double helix below ambient temperature. This occurs because of the high stability of complexes formed

between these proteins and single-stranded regions of DNA. Therefore the association of binding proteins with transient single-stranded regions, formed by the prior action of helicase, can displace the equilibrium between the double helix and the single-stranded forms in favor of the latter.

Topoisomerases. Even after the local unwinding and the separation of the strands is achieved, other practical problems must still be solved in order for the replication to proceed unimpeded. For the *E. coli* DNA it may be calculated, based on the rate at which a fork is propagated, that the parental double helix must unwind at a rate of at least 4,000 turns per minute. It is apparent that these high rates would generate serious practical difficulties if strands were to separate over an appreciable length of DNA. The large free energy requirements of bringing about the unwinding of large regions of DNA can, however, be reduced to manageable levels by the nicking of one of the DNA strands near the replicating fork. Since the fork is a moving entity, the nicking must be visualized as a reversible cut-and-rejoin process, which moves along with the fork.

Nicking is indispensable for a topological reason as well. Unwinding at one of the two forks requires that the parental double helix rotates in the opposite direction to that necessary for the unwinding of the opposite fork. Furthermore, even if only one of the forks is considered and the other is ignored, the topological problem still remains unresolved. In the absence of a nick as the unwinding at one of the forks would progress, an increasing number of positive supercoils would have to be introduced into the double helix. Once the limit of the helix to accommodate the supercoils was reached, the unwinding and the replication would have to stop.

The above topological restraints can be overcome if "mobile swivels" operate near the moving forks. Such swivels could originate from the transient breaking of a single phosphodiester bond in the parental double helix by the action of an endonuclease. The nick near the moving fork could then be resealed by a ligase. In this manner specialized enzymes with combined hydrolytic and ligating activities, the *topoisomerases,* may prevent DNA from forming positive superhelical turns as the synthesis progresses.

One of the topoisomerases isolated from *E. coli, gyrase,* has been found to induce the formation of negative supercoils in the presence of ATP, but in the absence of ATP it can also remove from DNA any supercoils, either positive or negative. These operations involve the breaking of the DNA backbone and the simultaneous formation of a DNA-gyrase covalent bond. The subsequent resealing of the severed phosphodiester bond occurs when the polynucleotide–gyrase complex transfers its polynucleotide constituent so that it can be

joined with the remaining polynucleotide from which it was severed. Gyrase thus acts by cleaving and resealing a phosphodiester bond in one of the parental strands, while leaving the other strand intact.

It is possible that one of the in vivo functions of gyrase is to remove the positive superhelical turns introduced into double-stranded DNA during replication. In addition to providing a swivel, the enzyme may facilitate replication by keeping the replicating molecule under negative superhelical stress, which favors the local separations of the strands at the expense of ATP.

A Model for the Replication of DNA

Our understanding of the mechanism of DNA replication is still far from complete. Extensive studies in *E. coli* and its phages have permitted the proposal of a replication model for this system. With the specific exceptions noted in the preceding sections regarding the substeps of the replication process, similar models may be applicable to mammalian systems.

The synthesis of *E. coli* DNA begins at the origin of replication of the chromosome. The very first strand synthesized the leading strand, may in this instance be synthesized in a continuous manner (Figure 17.56). As the leading strand grows, a fork is formed by the action of helicase. The strain that develops as a result of the unwinding of the strands within the parental closed circular DNA may be

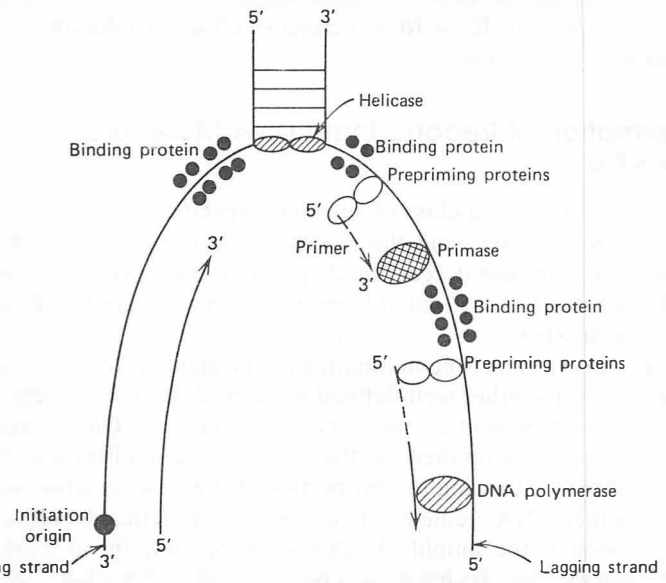


Figure 17.56
***Model for DNA replication in* E. coli.**
In this figure the initial stages of replication are depicted. The primers are subsequently removed from the newly synthesized segments of DNA at the lagging strand, and the segments are joined. Since replication is normally bidirectional, similar events take place concurrently at the other side of the initiation origin.

released by gyrase, which may also facilitate the separation of the strands by keeping the circular DNA under negative superhelical stress. Binding protein stabilizes the single-stranded portions of the unwound double helix so that the single strands are available for template function. Attachment of a prepriming protein on the single-stranded portion of the opposite strand sets the stage for the discontinuous synthesis of the lagging strand. Primase, the action of which is triggered by the prepriming proteins, synthesizes a brief complementary segment of the strand. This segment serves as a primer for covalent extension of the strand by DNA polymerase III and for the formation of Okazaki fragments. Removal of the primer portions at the 5′ end of the Okazaki fragments by DNA polymerase I, repair by the same enzyme, and joining of the repaired fragments by DNA ligase produce intact DNA strands.

17.8 RECOMBINANT DNA

The extent to which DNA from different organisms "mixes" in nature is strictly controlled, with combinations occurring only between closely related DNAs. The most common example of DNA "mixing" is the *integration* of phage or plasmid DNA into the corresponding bacterial hosts. Another form of DNA exchange occurs as certain phages incorporate small segments of DNA of the host bacterium and transfer it to a recipient cell upon infection, a process termed *transduction*.

Formation of Recombinant DNA Molecules In Vitro

The discovery of a class of sequence-specific nucleases, the *restriction endonucleases,* and the development of appropriate new techniques has in recent years made possible the in vitro recombination of DNA molecules isolated from organisms that are totally unrelated to one another.

The goal of most recombination experiments is to separate a single gene, or some other well-defined portion of DNA, from other genetic information present in the genome of the cell. Once separated, a gene must be amplified so that sufficient quantities can be made available for studying the properties of the gene in vitro, separately from other DNA elements. There are also practical benefits that can be gained if the amplified (cloned) gene can, in addition, be expressed in vitro. Technology is now available that allows the implan-

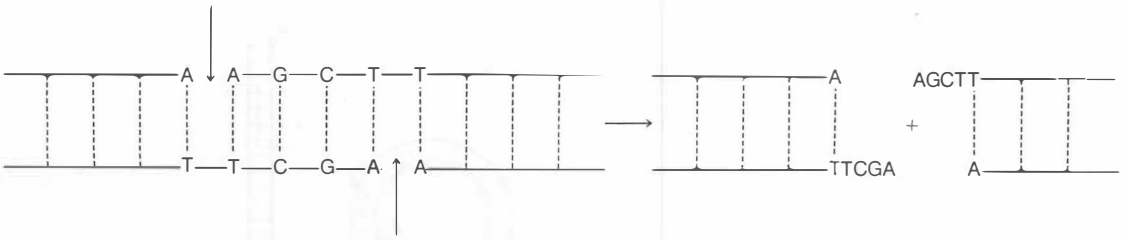


Figure 17.57
Sites of cleavage of DNA by the restriction enzyme Hind III.

tation of an intact piece of genetic information isolated from one organism, or even chemically synthesized, into another host organism where the implanted DNA can be both replicated and expressed.

Recombinant DNA technology is based on the use of restriction enzymes for cutting out well-defined segments of DNA from each of two chromosomes to be used in recombination. The recombination of DNA segments from different sources is very much facilitated because of the unusual manner in which most restriction enzymes introduce cuts into the DNA strands. Specifically, the cleavage of a palindromic sequence usually occurs at different positions along the two complementary strands, leaving two short polynucleotide stubs that are the antiparallel complements of one another. For the enzyme *Hind III*, isolated from a strain of *Haemophilus influenzae*, these points are indicated in Figure 17.57 by two arrows. The use of the same restriction enzyme for cutting *two different DNA molecules* produces segments with homologous terminal sequences. This homology permits a close fitting of the two segments, which is essential for recombination.

In the example shown in Figure 17.58 a small circular DNA isolated from *E. coli*, known as a *pBR322* plasmid, is used as the primary source (cloning vector or vehicle) of DNA. This plasmid, which has only a single palindrome susceptible to hydrolysis by Hind III, is cut at two points within the palindrome. Therefore only one double-stranded DNA molecule, with terminals containing two antiparallel single-stranded portions, AGCT and TCGA, is produced. The DNA to be inserted into the plasmid (foreign DNA or donor DNA) is also cut by Hind III, generating one or more segments, depending on how many appropriate palindromes it contains. These segments are destined to terminate in short single-stranded stubs identical to those present in the severed plasmid.

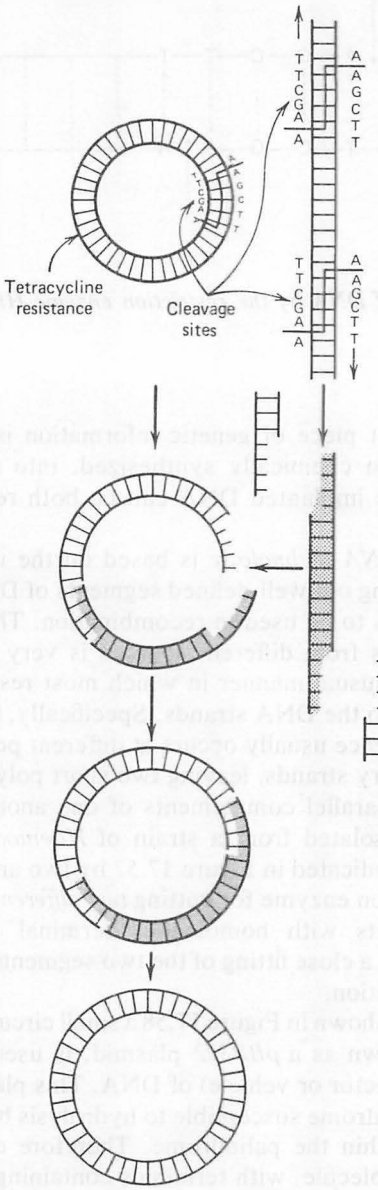


Figure 17.58
Formation of recombinant DNA.
In constructing a recombinant DNA molecule, foreign DNA is joined to a cloning vector and the recombinant is introduced into a bacterial cell for amplification. The basic steps involved are as follows. (a) The cloning vector and the foreign DNA are cleaved at homologous sites (i.e., at sites with the same nucleotide sequence) by a restriction enzyme, often selected so that it cleaves the cloning vector at a single site and the donor DNA at several sites. (b) In the sequences that are cleaved the enzyme creates staggered cuts that are homologous between the cloning vector and the donor DNA. Under annealing conditions an opened cloning vector will add a fragment of foreign DNA, and then the circle will close. The nicks in the annealed ends are then sealed by the action of ligase, which forms a stable circular DNA. (c) The recombinant DNA is taken up by a bacterial cell and replicated during cell division.

Annealing between the segments generated by separate treatment of the two different DNAs produces a number of new DNA combinations. The key to the success of the splicing of the donor DNA into the vector is the presence, at the end of each fragment, of the complementary single-stranded polynucleotide stubs generated by digestion with Hind III or some other restriction nuclease. Treatment with ligase seals the open phosphodiester bonds and es-

tablishes a number of different recombinant DNA molecules. Among the combinations are circular DNA molecules consisting of the plasmid into which one *appropriate* segment of donor DNA has been inserted.

Cloning of Recombinant DNA Molecules

The method used for recombining different DNAs in the above example has almost unlimited applicability because it places few restrictions in the selection of the DNAs that can be used for the recombination. In fact, the only apparent limitation for the recombination is the existence of identical palindromes among the sequences of DNAs to be used for splicing.

In practice, however, the formation of a few recombinant molecules falls short of the goal of producing sufficient quantities of recombinant DNA with specific composition necessary for most applications. What is needed, in addition, is a method that permits the *selection* of the few molecules that contain the appropriate combinations of DNA fragments. Recombinant molecules should also be able to provide for their own replication so that their number can be amplified. The recombinant DNA molecules must therefore contain elements that allow for both their selection and amplification (cloning).

Although various types of DNA molecules could in principle be used as cloning vectors, plasmids are generally the most suitable and the most common vectors presently in use. These small extrachromosomal DNAs present in many bacteria can be used to advantage because they permit both the convenient selection and the cloning of recombinant DNA molecules. Selection is made by virtue of the fact that plasmids are the elements that convey bacterial resistance to antibiotics, that is, they contain genes that code for proteins that can degrade antibiotics. For this reason, a plasmid that confers resistance to the antibiotic tetracycline, for example, can provide a phenotype permitting the selection of only those cells that contain the plasmid. This can be achieved by simply growing the recombinant molecules in the presence of the antibiotic. In addition, plasmids, which replicate autonomously within replicating bacterial cells, generally do not lose their ability to replicate if segments of a donor DNA are incorporated into their chromosomes; therefore plasmids are suitable for amplifying the amounts of the recombinant moiety.

Selection of Appropriate Recombinant Molecules

Selection of only the desired recombinant molecules must discriminate for *both* segments of DNA present in the recombinant entity,

CLIN. CORR. 17.3 RECOMBINANT DNA

Recombinant DNA techniques are extensively used today. The reason why these techniques are receiving wide application is that they offer great promise both in basic research and for practical benefit. The complexity of eucaryotic genomes makes the study of a given gene and its expression extremely difficult. Recombination techniques now allow the study of an isolated (and amplified) gene by transplantation into a bacterial system. The possibilities for practical benefit are numerous. For instance, the insertion of a synthetic gene coding for the mammalian peptide hormone somatostatin into bacteria has induced bacteria to produce the hormone, that is, the synthetic gene can be replicated, transcribed, and translated by the host. Similarly, the gene coding for mammalian proinsulin has been appropriately recombined and grown in *E. coli*. As a result, bacteria can be made to secrete proinsulin. Numerous other achievements in this area are presently taking place at a high rate. In fact, the potential for benefit from recombinant DNA techniques appears enormous. This technology may be expected to increase the supply of hormones and other very rare proteins, for example, interferon, which show great promise as new tools in medicine.

that is, in addition to selecting DNA molecules that contain the plasmid the selection must be further narrowed to only those recombinant molecules that have incorporated the appropriate segment of donor DNA. The additional selection requirement can be easily satisfied in the simple case when the donor DNA happens to be also a segment of a plasmid containing its own "marker," for example, the gene responsible for penicillin resistance. In such a case the ligated mixture of recombinant DNA molecules, some of which contain both the vector (plasmid 1) DNA and a segment of the donor (plasmid 2) DNA, can be added to recipient *E. coli* cells. The recipient cells can then be cultivated in a medium containing both tetracycline and penicillin. Obviously, only the recombinant plasmids resistant to both antibiotics will grow. In other instances selection for the donor gene can be accomplished if the gene can be expressed in the *E. coli* host. For example, the gene that codes for one of the enzymes of histidine biosynthesis, imidazole glycerol phosphate dehydratase, in *yeast*, has been isolated by growing the recombinant DNA molecules into an appropriate histidine auxotroph and selecting for cells that do not require histidine for growth.

"Marking" for the "donor" DNA molecules is not always as simple. When convenient selection procedures such as those mentioned above cannot be used, other methods must be applied to "fish out," from a culture containing vector molecules as well as vector–donor combinations, only the desired vector–donor combinations. For instance, whenever a polynucleotide with nucleotide sequences complementary to those of the donor DNA is available, hybridization techniques can be used to select for the donor DNA. This method has been used for the selection of recombinant bacterial DNA molecules containing the 18S and 28S RNA genes of frog (*X. laevis*) cells. In this instance the DNA transcripts of these genes were used for hybridization.

Clin. Corr. 17.3 discusses applications of recombinant DNA technology.

17.9 NUCLEOTIDE SEQUENCING OF DNA

Although the first determination of the primary structure of a small RNA molecule, for example, tRNA, occurred in the 1960s, the complete sequencing of larger RNA and DNA molecules remained until the mid-1970s an enormously laborious undertaking. The de-

termination of the sequence of a large DNA molecule often begins by cutting the DNA, by the use of appropriate restriction endonucleases, into pieces of a more manageable size. In recent years several effective methods have been developed for the rapid sequencing of relatively large polynucleotides; the nucleotide sequences of several viral genomes have been determined using these methods. Clearly, the determination of the sequences of a double-stranded DNA is, in principle, an undertaking equivalent to establishing the sequence of only one of the two strands. The sequence of the other strand is obviously fixed by the requirement for complementarity between strands. In practice, however, the results provided by sequencing one polynucleotide are not always completely unequivocal and are often supplemented by determining the sequence of the complementary strand as well.

"Plus and Minus" Method

The first procedure developed for sequencing large polynucleotides is known as the "plus and minus" method. This method, and a modification of this method outlined in this section, *do not determine the nucleotide sequence of a polynucleotide directly*. Instead, the *terminal nucleotides* of a large number of polynucleotide segments of various sizes, smaller than the polynucleotide undergoing sequencing, are determined. Each of these segments is complementary to gradually increasing portions of the polynucleotide that is being sequenced, in that they are variable length transcripts (Figure 17.59). As these *transcripts* are synthesized by DNA polymerase, a primer is

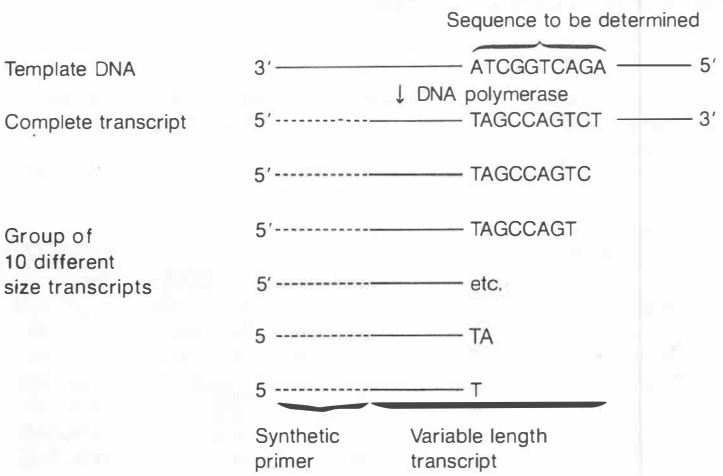


Figure 17.59
Synthesis of variable length transcripts for determination of a nucleotide sequence in DNA. *In this example a hypothetical 10 nucleotide sequence is transcribed in a manner that produces a series of variable length transcripts, 10 in all, which terminate at various nucleotides within the sequence to be determined. Use of a synthetic primer insures that all transcripts are "in register" relative to the sequence of interest.*

clearly needed, which is normally obtained by separate digestion of the double-stranded DNA undergoing sequencing, using two endonucleases. Suitable choice of these nucleases allows the separation of a short primer, which can associate at the left side of the 5′ terminal of the polynucleotide region to be sequenced. Alternatively, when the sequence with which the primer must associate is known, a synthetic polynucleotide can be used as a primer. The primer is then elongated as usual, by the stepwise incorporation of nucleotides complementary to those present in the template.

The detailed experimental approaches used in the original plus and minus method will not be cited here. Instead, a newer simplified and more accurate sequencing approach, based on the same principles as used in the plus and minus method, is described below. Both the original plus and minus method and this newer sequencing method are based on the use of DNA polymerase for transcribing the DNA regions to be sequenced under controlled conditions. Each of the transcripts is thus forced to terminate in a preselected base. The lengths of various families of transcripts obtained are used for determining the relative position of their terminal nucleotides in the template undergoing sequencing.

"Forced Termination" Method

The plus and minus method depends on the DNA polymerase catalyzed synthesis of "variable length transcripts" under four different sets of conditions. For each set, one of the four deoxyribonucleoside triphosphates is omitted from the replicating system. For this reason the corresponding transcripts progress only up to a point and terminate at the base just preceding the one omitted. This is detailed in the example depicted in Figure 17.60.

Complete set of variable length transcripts of the sequence to be determined	Complementary segments obtained in the absence of			
	dATP	dGTP	dCTP	dTTP
10 5′-TAGCCAGTCT 3′	—	—	—	—
9 -TAGCCAGTC	—	—	—	-TAGCCAGTC
8 -TAGCCAGT	—	—	-TAGCCAGT	-TAGCCAGTC
7 -TAGCCAG	—	—	-TAGCCAGT	-TAGCCAG
6 -TAGCCA	—	-TAGCCA	-TAGCCAGT	-TAGCCAG
5 -TAGCC	-TAGCC	-TAGCCA	-TAGCCAGT	-TAGCCAG
4 -TAGC	-TAGCC	-TAGCCA	-TAGC	-TAGCCAG
3 -TAG	-TAGCC	-TAGCCA	-TAG	-TAGCCAG
2 -TA	-TAGCC	-TA	-TAG	-TAGCCAG
1 -T	-T	-TA	-TAG	-TAGCCAG

Figure 17.60
Elongation of variable length transcripts in the absence of one type of nucleotide.
The variable length transcripts of a DNA segment undergoing sequencing are elongated by DNA polymerase. Four sets of elongation products can be produced. Each one of these sets is obtained in the absence of one of the four nucleotides from the polymerizing mixture. The products in each set depend on the missing nucleotide. For instance, in the absence of deoxy adenosine triphosphate (column 1) the transcription terminates either just before the first A, giving as transcript -T (last line), or if the first A is already in place it will continue until just before the second A (lines 2 to 5), giving -TAGCC. Finally, if the second A is already in place, the transcription continues beyond the sequence of interest (lines 6 to 10). Similar considerations apply to the other three sets obtained in the absence of guanosine, cytosine, and thymidine nucleotides (columns 2 to 4), respectively. Boxed transcripts identify the distinct oligonucleotides in each column. These polynucleotides are identified by electrophoresis.

In the newer "forced termination" method all four deoxyribonucleoside triphosphates are present, but in each one of the four sets a structural analog of one of the bases is also added. For instance, in one of the sets the transcription of the strand to be sequenced is carried out in the presence of the usual four nucleotides but with 2′,3′-dideoxythymidine triphosphate (ddTTP) also present. DNA polymerase incorporates ddTTP in place of deoxyribonucleotide triphosphates (dTTP) to an extent that depends on the ratio between ddTTP and dTTP. The polynucleotide chain cannot be extended beyond thymine positions into which, instead of dT, a ddT molecule has been incorporated because ddTTP has no 3′-hydroxyl group (Figure 17.61). Under these conditions a mixture of fragments of various lengths is obtained, all having the same 5′ terminus and with ddT residues at the 3′ ends. This mixture can be fractionated by electrophoresis and the pattern of bands reveals the distribution of dTs in the transcripts. Analogous terminators are used for the other nucleotides in separate experiments. All the transcripts are then run in parallel on an electrophoresis gel, and from the pattern of bands obtained the polynucleotide sequences can be read off.

The advantages of this method over the original plus and minus method are that it gives more clear-cut results and permits longer stretches of polynucleotide sequences to be read from a single gel. It is also simpler to carry out. In general, up to 200 nucleotides from the priming site can be determined with reasonable accuracy by using a single transcript. The complete nucleotide sequence of DNA molecules containing several thousand bases, or DNA of any size, for that matter, can be determined by prior enzyme fragmentation of the molecule. Polynucleotides of sizes suitable for sequencing can be obtained and the complete sequence of the DNA determined.

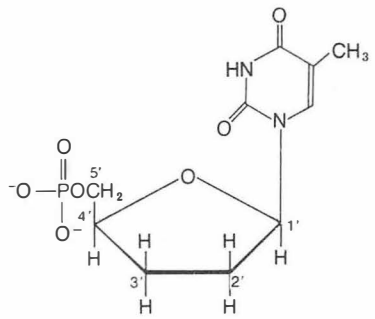


Figure 17.61
The structure of 2′,3′-dideoxythymidine.

BIBLIOGRAPHY

Books

Cantor, C. R., and Schimmel, P. R. *Biophysical chemistry,* Part I: *The conformation of biological macromolecules.* San Francisco: W. H. Freeman, 1980.

Cantor, C. R., and Schimmel, P. R. *Biophysical chemistry,* Part III: *The behavior of biological macromolecules.* San Francisco: W. H. Freeman, 1980.

Davidson, J. N. *The biochemistry of the nucleic acids.* New York: Academic, 1977.

Davidson, J. N., and Cohn, W. E. (eds.). *Progress in nucleic acid research and molecular biology*. New York: Academic (A series published regularly).

Freifelder, D. *The DNA molecule: structure and properties*. San Francisco: W. H. Freeman, 1978.

Guschlbauer, W. *Nucleic acid structure*. Berlin: Springer-Verlag, 1976.

Kornberg, A. *DNA replication*. San Francisco: W. H. Freeman, 1980.

Watson, J. D. *Molecular biology of the gene*, 3rd ed. Menlo Park, Calif.: W. A. Benjamin, 1976.

Woods, R. A. *Biochemical genetics*. New York: Chapman and Hall, 1980.

Articles

Champoux, J. J. Proteins that affect DNA conformation. *Annu. Rev. Biochem.* 47:449, 1978.

Cozzarelli, N. R. DNA gyrase and the supercoiling of DNA. *Science* 207:953, 1980.

Crick, F. Split genes and RNA splicing. *Science* 204:264, 1979.

Davis, B. D. The recombinant DNA scenarios: andromeda strain, chimera and golem. *Am. Sci.* 65:548, 1977.

Olins, D. E., and Olins, A. L. Nucleosome: the structural quantum in chromosomes. *Am. Sci.* 66:704, 1978.

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463, 1977.

Seltow, R. B., Repair deficient human disorders and cancer. *Nature* 271:713, 1978.

Smith, H. O. Nucleotide sequence specificity of restriction endonuclease. *Science* 205:455, 1979.

18

RNA

PETER N. GRAY

TRANSCRIPTION AND POSTTRANSCRIPTIONAL MODIFICATION

18.1 OVERVIEW

The primary genetic information of a eucaryotic cell is localized within the cell nucleus and encoded in the DNA. Expression of part or all of this information, a cell's functional capacity, or phenotype, usually takes the form of polypeptides, which are synthesized in the cytoplasm. Since genomic DNA is not a normal cytoplasmic component, cells require mechanisms for transferring the necessary information from the nucleus to the cytoplasmic protein-synthesizing machinery. Macromolecules that mediate the transfer of information must reflect the sequence of the purines and pyrimidines of the DNA. These macromolecules, called ribonucleic acids (RNAs), are linear polymers of ribonucleoside 5′-monophosphates. The process by which the RNA copies of selected DNA sequences are made is termed *transcription*.

Analysis of total cellular RNA reveals several distinct molecular sizes and functional families of RNA. These are presented in Table 18.1. The primary role of RNA within the cell is its involvement in protein synthesis, where messenger RNAs serve as templates, transfer RNAs as amino acid carriers, and ribosomal RNAs function in ribosomes during peptide bond formation. RNA involvement is required, since there is no specific affinity for amino acids by the purines and pyrimidines. Cytoplasmic RNA molecules, which contain the information for directing synthesis of a specific polypeptide, are messenger RNAs (mRNA). The size of each mRNA is directly related to the size of the protein for which it codes. The molecules that transfer specific amino acids from soluble amino acid pools to ribosomes, and ensure the alignment of these amino acids in the proper sequence prior to peptide bond formation, are transfer RNAs (tRNA). All the tRNA molecules are approximately the same size and shape. The assembly site, or factory, for peptide synthesis involves ribosomes, complex subcellular particles containing at least three different RNA molecules called ribosomal RNAs (rRNA), and 70 to 80 ribosomal proteins.

Protein synthesis requires a close interdependent relationship between mRNA, the informational template, tRNA, the amino acid adaptor molecule, and rRNA, part of the synthetic machinery. In order for protein synthesis to occur at the correct time in a cell's life, the synthesis of mRNA, tRNA, and rRNA must be coordinated with the cell's response to the intra- and extracellular environments.

All cellular RNA is synthesized on a DNA template and reflects a portion of the DNA base sequence. Therefore, all RNA is associated with DNA at some time. The RNA involved in protein synthesis

Table 18.1 Characteristics of Cellular RNAs

Type of RNA	*Abbreviation*	*Function*	*Size and Sedimentation Coefficient*	*Site of Synthesis*	*Structural Features*
Messenger RNA		Transfer of genetic information from nucleus to cytoplasm, or from gene to ribosome	Depends upon size of protein		Blocked 5′ end; poly(A) tail on 3′ end; nontranslated sequences before and after coding regions; few base pairs and methylations
Cytoplasmic	mRNA		1,000 to 10,000 nucleotides	Nucleoplasm	
Mitochondrial	mt mRNA		9S to 40S	Mitochondria	
Transfer RNA					
Cytoplasmic	tRNA	Transfer of amino acids to mRNA · ribosome complex and correct sequence insertion	65–110 nucleotides 4S	Nucleoplasm	Highly base-paired; many modified nucleosides; common specific structure
Mitochondrial	mt tRNA		3.2S to 4S	Mitochondria	
Ribosomal RNA					
Cytoplasmic	rRNA	Structural framework for ribosomes	28S, 5,400 nucleotides	Nucleolus	5.8S and 5S highly base-paired; 28S and 18S have some base-paired regions and some methylated nucleosides
			18S, 2,100 nucleotides	Nucleolus	
			5.8S, 158 nucleotides	Nucleolus	
			5S, 120 nucleotides	Nucleoplasm	
Mitochondrial	mt rRNA		16S, 1,650 nucleotides	Mitochondria	
			12S, 1,100 nucleotides		
Heterogeneous nuclear RNA	HnRNA	Some are precursors to mRNA and other RNAs	Extremely variable 30S to 100S	Nucleoplasm	mRNA precursors may have blocked 5′ ends and 3′-poly(A) tails; many have base-paired loops
Small nuclear RNA	snRNA	Structural and regulatory RNAs in chromatin	100–300 nucleotides	Nucleoplasm	

functions in the cytoplasm outside the nucleus, while some RNA remains in the nucleus, where they have structural and regulatory roles. RNA synthesized in the mitochondria remains there and is involved in mitochondrial protein synthesis.

Although DNA is the more prevalent genetic store of information, RNA can also carry genetic information in the sequence of the bases and serves as the genome in several viruses. However, RNA is not normally found as the genome in eucaryotic or procaryotic cells. Genomic RNA is found in the RNA tumor viruses and the other small RNA viruses, such as poliovirus and reovirus.

18.2 STRUCTURE OF RNA

Components and Primary Structure

RNA has been shown to be a general constituent of eucaryotes and procaryotes. Chemically it is very similar to DNA, but unlike DNA, RNA is not very stable within a cell. It usually is synthesized, used, and rapidly degraded. RNA is an unbranched linear polymer in which the monomeric subunits are the ribonucleoside 5′-monophosphates.

The purines found in RNA are *adenine* and *guanine;* the pyrimidines are *cytosine* and *uracil.* Except for uracil, which replaces thymine, these are the same bases found in DNA. Complete analysis of cellular RNA reveals that other bases are also found in low concentrations. These minor bases represent modifications of the basic ring structures in A, G, U, and C and are a small but significant portion of bases having prominent roles in RNA metabolism and function (Table 18.2). Modification of the purines and pyrimidines occurs only after polymerization of the nucleotides. Modifications to the ribose, usually methylation of the 2′-hydroxyl position, also occur after polymerization.

The monomers are connected by single phosphate groups linking the 5′-carbon of one ribose with the 3′-carbon of the next ribose. This internucleotide link, a 3′,5′ phosphodiester, forms a chain or backbone from which the bases extend like the side groups of the amino acids in proteins (Figure 18.1). The length of natural RNA molecules in eucaryotic cells varies from approximately 65 nucleotides to more than 6,000 nucleotides. The sequences of the bases are complementary to the base sequences of specific portions of only *one strand of DNA.* Thus, unlike the base composition of DNA, molar ratios of A + U and G + C in RNA are not equal. All cellular RNA so far examined is linear and single-stranded, but double-stranded RNA is present in some viruses.

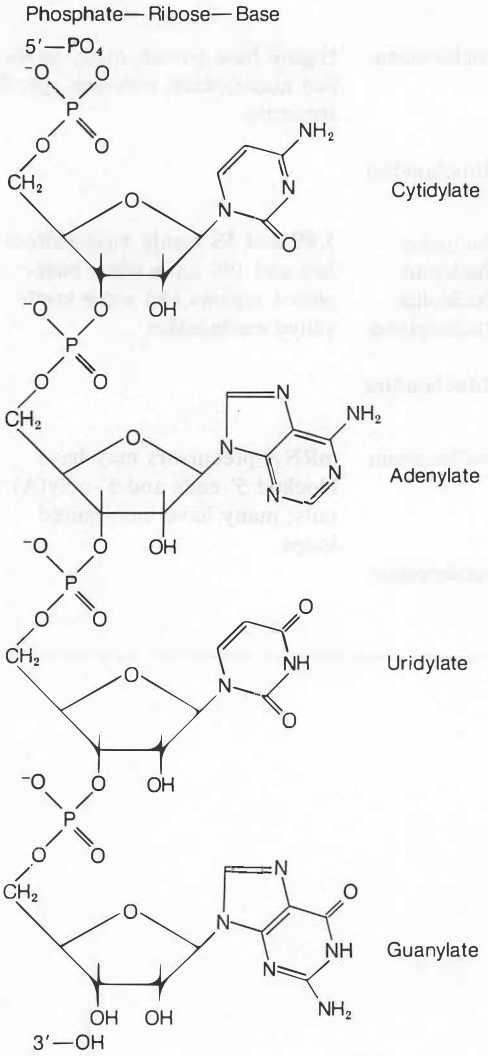


Figure 18.1
The structure of the 3′,5′ phosphodiester bonds between ribonucleotides forming a single strand of RNA.
The phosphate joins the 3′-OH of one ribose with the 5′-OH of the next ribose. This linkage produces a polyribonucleotide having a sugar-phosphate "backbone." The purine and pyrimidine bases extend away from the axis of the backbone and may pair with complementary bases to form double-helical base-paired regions.

Table 18.2 Some Modified Nucleosides Found in RNA

Purine Derivatives

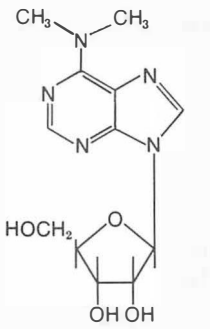
Nucleosides with a methylated base
- 1-Methyladenosine (m^1A)
- 2-Methyladenosine (M^2A)
- 7-Methylguanosine (m^7G)
- 1-Methylguanosine (m^1G)
- N^6-Methyladenosine (m^6A)
- N^6,N^6-Dimethyladenosine (m_2^6A)
- N^2-Methylguanosine (m^2G)
- N^2,N^2-Dimethylguanosine (m_2^2G)

2′-*O*-Methylated derivatives
- 2′-*O*-Methyladenosine (Am)
- 2′-*O*-Methylguanosine (Gm)

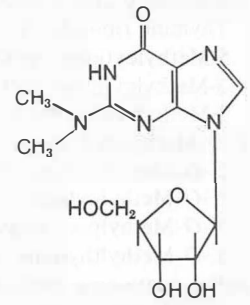
Deaminated derivatives
- Inosine (I)
- 1-Methylinosine (m^1I)

Adenosine derivatives with an isopentenyl group
- N^6-(Δ^2-Isopentenyl)adenosine (i^6A)
- N^6-(Δ^2-Isopentenyl)-2-methylthioadenosine (ms^2i^6A)
- N^6-(4-Hydroxy-3-methylbut-2-enyl)-2-methylthioadenosine

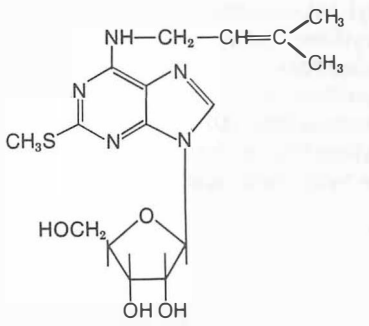
Other nucleosides
- *N*-[9-(β-D-Ribofuranosyl)purin-6-ylcarbamoyl]threonine (t^6A)
- *N*-[9-(β-D-Ribofuranosyl)purin-6-yl-*N*-methylcarbamoyl]threonine (mt^6A)



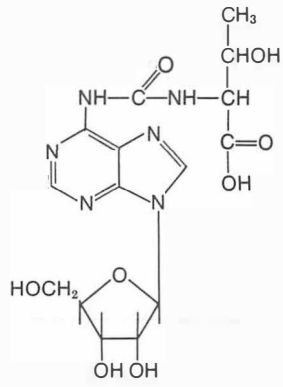
N^6,N^6-Dimethyladenosine



N^2,N^2-Dimethylguanosine



N^6-(Δ^2-Isopentenyl)-2-methylthioadenosine



N-[9-(β-D-Ribofuranosyl)purin-6-yl carbamoyl]-L-Threonine

(Continued on p. 876)

Table 18.2 Some Modified Nucleosides Found in RNA (*Continued*)

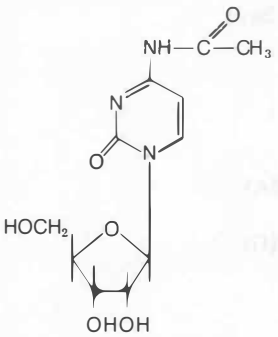
Pyrimidine Derivatives

Nucleosides with a methylated base
- Thymine riboside (T)
- 5-Methylcytidine (m^5C)
- 3-Methylcytidine (m^3C)
- 3-Methyluridine (m^3U)

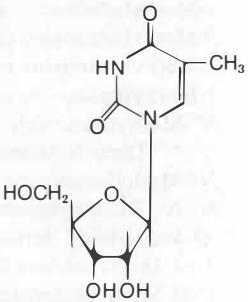
2′-*O*-Methylated derivatives
- 2′-*O*-Methyluridine (Um)
- 2′-*O*-Methylcytidine (Cm)
- 2′-*O*-Methylpseudouridine (Ψm)
- 2′-*O*-Methylthymine riboside (Tm)

Sulfur-containing nucleosides
- 4-Thiouridine (s^4U)
- 5-Carboxymethyl-2-thiouridine methyl ester
- 5-Methylaminomethyl-2-thiouridine
- 5-Methyl-2-thiouridine
- 2-Thiocytidine

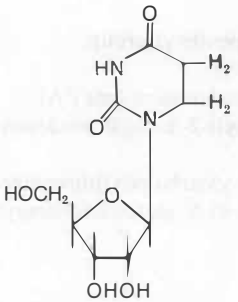
Other nucleosides
- Pseudouridine (Ψ)
- 5,6-Dihydrouridine (D)
- 4-Acetylcytidine (ac^4C)
- Uridine-5-oxyacetic acid



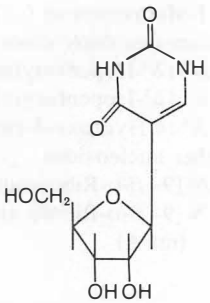
*N*⁴-Acetylcytidine



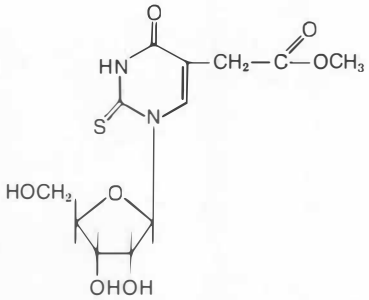
5-Methyluridine



5,6-Dihydrouridine



5-(β-D-Ribofuranosyl)uracil
(Pseudouridine)



2-Thio-5-carboxymethyluridine methyl ester

Secondary Structure of RNA

RNA in solution exists in a partially ordered configuration. In low ionic strength solutions the molecules appear as extended polyelectrolyte chains. Shifting the molecules to solutions of high ionic strength causes the RNA to contract. Increasing the temperature denatures the RNA by disrupting the hydrogen-bonded base pairs and the base stacking. These changes can be monitored by measuring absorption of ultraviolet light at 260 nm, much like DNA. Since the RNA is single-stranded, the hyperchromic shifts from an ordered molecular conformation, a partial helix, to an extended and then random coil are not as high as for DNA. Considerable helical structure exists in RNA in the absence of extensive base pairing. This helix is due to the strong base-stacking forces between A, G, and C residues. Base stacking is more important than simple hydrogen bonding in determining how nucleic acids interact with each other.

These forces act to restrict the conformation of RNA (Figure 18.2). Base stacking is a result of the van der Waals forces between the π electron clouds above and below the unsaturated rings of the purines and pyrimidines and the hydrophobic nature of the bases. The distance restriction of the phosphodiester-ribosyl backbone and the near-perpendicular angle of the β-glycosidic bond do not permit the bases to stack directly over each other. Therefore each succeeding base is offset by ~35°, forming a helical structure with 10 or 11 nucleotides per turn in a double helix and offset ~60° in a single strand of RNA, producing six nucleotides per turn.

A single RNA molecule may also have double-helical regions formed by hydrogen bonding between complementary base sequences located within the molecule, as occurs in DNA. These duplex structures, often called "hairpins", may or may not have large unpaired loops at the end. There are considerable variations in the fine structural details of the "hairpin" structures. These variables include the length of base-paired regions and the size and number of unpaired loops (Figure 18.3). tRNAs have a large proportion of their bases involved in these helical structures and are excellent examples of base stacking and hydrogen bonding in a single-stranded molecule (Figure 18.4*a*). The anticodon region in tRNA is an unpaired loop of seven bases. The partial helix caused by base stacking in this loop binds, by specific base pairing, to a complementary codon in mRNA so that translation (peptide bond formation) can occur. Within the tRNA molecule itself ~60% of the bases are paired in four double-helical regions called stems. In addition, the unpaired regions have the capability to form base pairs with free bases in the same or other looped regions. The role and extent of base pairing in each type of RNA is described in the following sections.

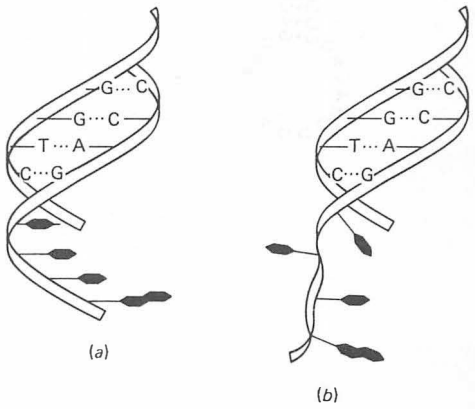


Figure 18.2
Models indicating a helical structure due to (a) base stacking in the —CCA_{OH} terminus of tRNA and (b) the lack of an ordered helix when no stacking occurs in this non-base-paired region.
Redrawn from M. Sprinzl and F. Cramer, *Prog. Nucl. Res. Mol. Biol.*, 22:9, 1979.

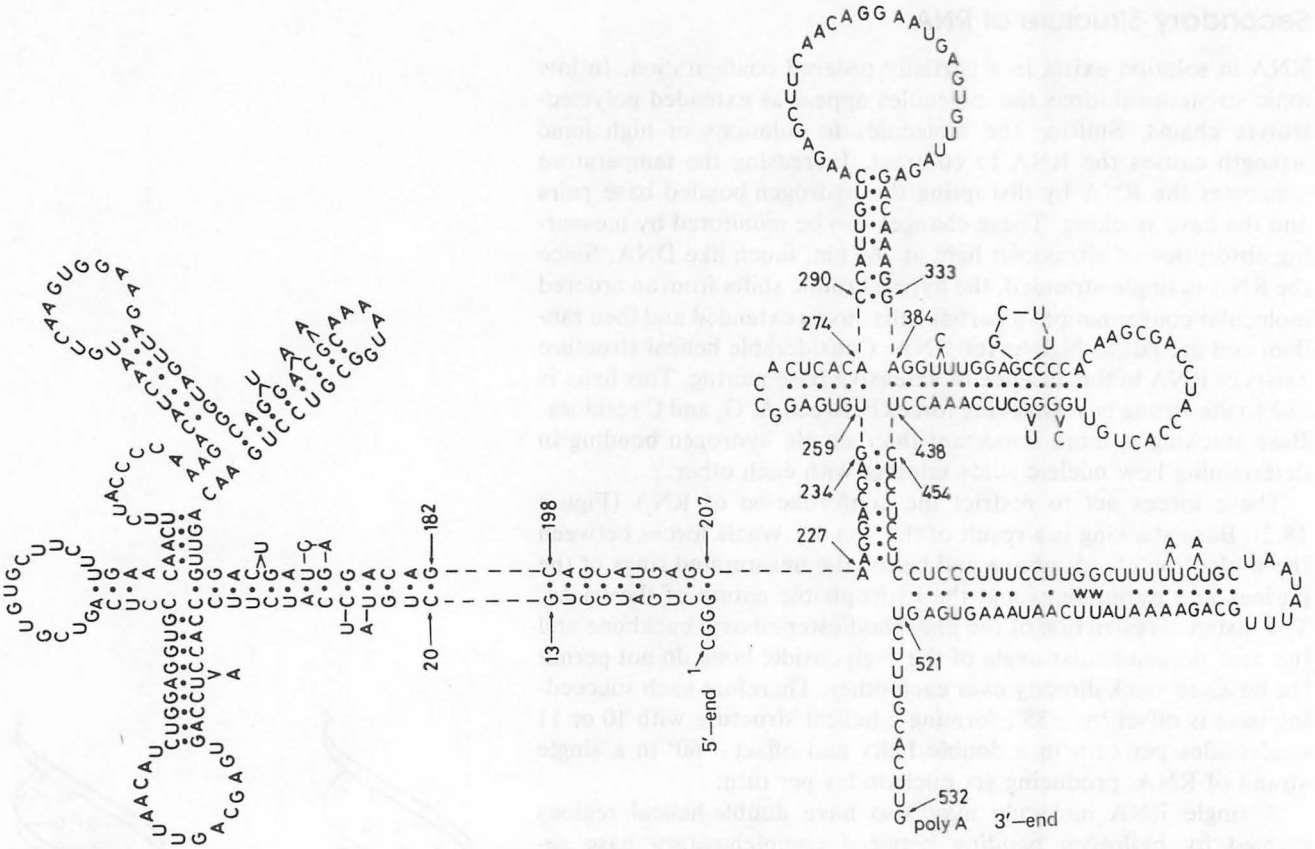


Figure 18.3
Proposed base-pairing regions in the mRNA for mouse immunoglobulin light chain.
Base-paired structures shown have free energies of at least −5 kcal. Note the variance in loop size and length of paired regions.
P. H. Hamlyn et al., *Cell*, 15:1073, 1978. Reproduced with permission.

Tertiary Structure of RNA

The actual functioning structures for the RNAs are more complex than the base-stacked and hydrogen-bonded helices mentioned above. The RNAs, in vivo, are dynamic molecules in solution, which undergo changes in configuration during synthesis, processing, and functioning. Like DNA, RNA is always associated with one or more proteins that have functional and conformational roles.

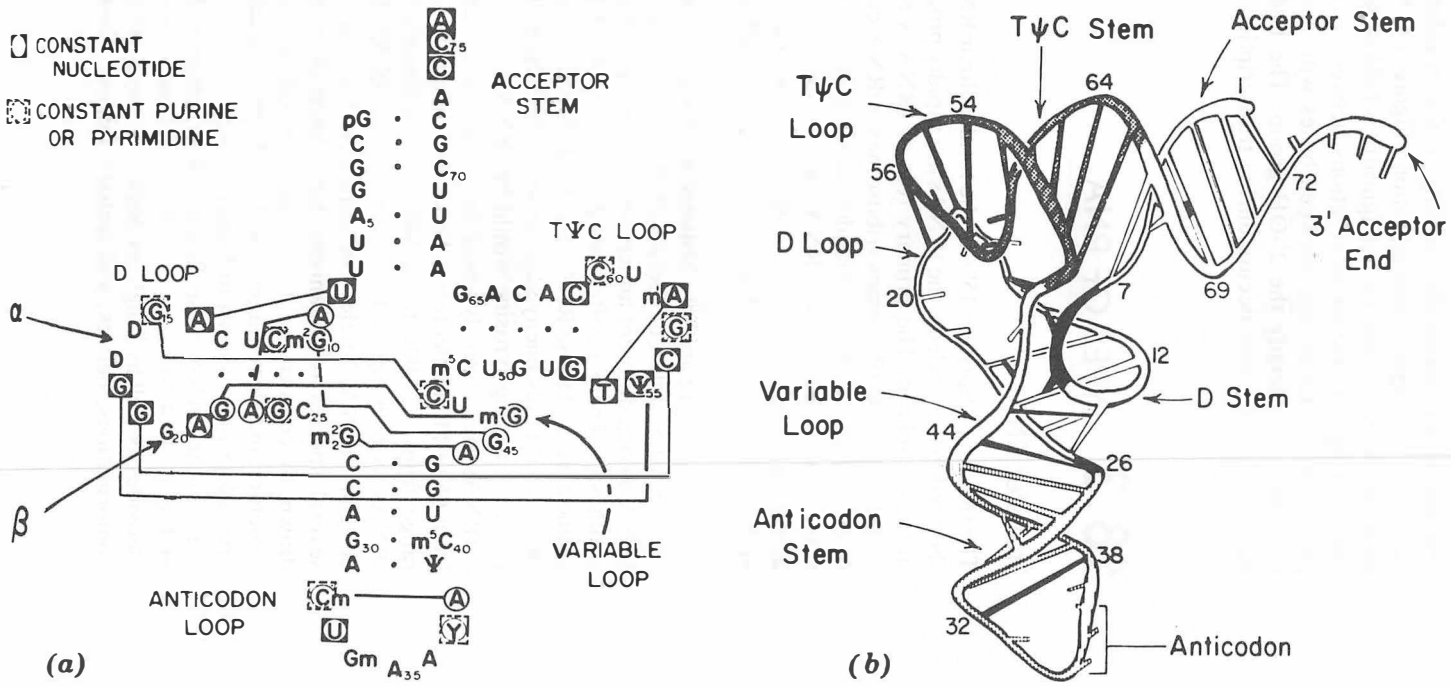


Figure 18.4

(a) Cloverleaf diagram of the two-dimensional structure and nucleotide sequence of yeast tRNA^{Phe}. Solid lines connecting circled nucleotides indicate hydrogen-bonded bases. Solid squares indicate constant nucleosides; dashed squares indicate a constant purine or pyrimidine. Insertion of nucleotides in the D loop occurs at positions α and β for different tRNAs. (b) Tertiary folding of the cloverleaf structure in (a). Hydrogen bonds are indicated by cross rungs.

These proteins lend stability to the RNA much like the base pairing and subsequent molecular folding. tRNA in solution is folded in a compact "L-shaped" configuration (Figure 18.4*b*). The arms and loops are folded in specific configurations held in position not only by base pairing but also by interactions between bases and the phosphodiester backbone and between bases with particular regions of the sugars, especially the 2′-OH group. The folding of the tRNA molecules apparently occurs during transcription.

18.3 TYPES OF RNA

Transfer RNA

The tRNAs comprise ~15% of the total cellular RNA. Although synthesized in the nucleus, the tRNAs are rapidly processed and utilized in the cytoplasm. The primary role of tRNA is to transport amino acids to the polyribosomes (ribosomes–mRNA complex) where the three nucleotides of the anticodon loop region bind to complementary nucleotide triplets of the mRNA. tRNAs therefore have two primary active sites, the $—CCA_{OH}$, *3′-hydroxyl terminus,* to which specific amino acids are enzymatically attached covalently, and the *anticodon triplet*.

Although there are only 20 amino acids used in most proteins, there are at least 56 different species of tRNAs in any given cell, with each tRNA having different anticodon triplets (see Chapter 19 on protein synthesis). There is often more than one tRNA for any particular amino acid, and these RNAs may be defined as isoacceptor tRNAs. A tRNA that binds phenylalanine would be written $tRNA^{Phe}$, whereas one for carrying tyrosine would be $tRNA^{Tyr}$.

tRNAs are relatively small for nucleic acids and range in length from 65 to 110 nucleotides. This corresponds to a molecular weight range from ~22,000 to 37,300. The sedimentation coefficient for tRNAs as a group is 4S, and the term 4S RNA is often used to designate tRNA. Nucleotide sequences for at least 85 tRNAs from several species of organisms have been determined and can be drawn to conform to the general two-dimensional "cloverleaf" structure and three-dimensional "L-form," as determined by x-ray crystallography, shown in Figure 18.4.

From the nucleotide sequence and structure of the $tRNA^{Phe}$ shown in Figure 18.4, it is clear that tRNAs have several modified nucleosides as well as a high proportion of bases involved in secondary conformations, helices, and tertiary folding. Some of the modified

nucleosides found in tRNA are listed in Table 18.2 and their positions in tRNA indicated in Figure 18.4. The modified nucleosides alter the structural stability of the RNA but are not essential for forming or maintaining a tertiary conformation. The modifications do not appear to have a role in general aminoacylation, or "charging," of tRNAs, but may be involved in specific and nonspecific recognition of enzymes and proteins and may specify interactions between tRNA and ribosomes.

Many structural features are common to all tRNA molecules (see Table 18.3). Seven base pairs are always in the amino acid acceptor stem, which, in functioning molecules, is terminated with the nucleotide triplet —CCA_{OH}. This —CCA_{OH} triplet is not base-paired. The dihydrouracil or "D" stem has three or four base pairs, while the anticodon and —TΨC— stems have five base pairs each. Both the anticodon loop and —TΨC— loop have seven nucleotides. Since the distance from the amino acid attachment to the anticodon triplet is constant, about 75 Å, the varying number of nucleotides in the different tRNAs are taken up by the variable loop. Thus 80% of tRNAs have small variable loops of 4 or 5 nucleotides, while the others have larger loops with 13 to 21 nucleotides. Five or six nucleotides may also be incorporated at two positions in the D loop, which varies from 8 to 14 nucleotides. Several of the nucleotides are in constant positions (see Figure 18.4*a*).

Ribosomal RNA

The eucaryotic cytoplasmic ribosomes are composed of four RNA molecules and 70 to 80 proteins. These RNAs and proteins are divided specifically between the two ribosomal subunits. The smaller subunit, the 40S particle, contains one 18S rRNA of ~2,100 nucleotides and 55% of the proteins. The large ribosomal subunit, the 60S particle, contains the remaining rRNAs and proteins. The three rRNAs in the large subunit are the 28S rRNA with 5,400 nucleotides, the 5.8S rRNA with 158 nucleotides, and the smaller 5S rRNA with 120 nucleotides. The approximate sedimentation coefficients and G + C contents are described in Tables 18.1 and 18.4.

The rRNAs account for 80% of the total cellular RNA and are metabolically stable. This stability, required for repeated functioning of the ribosome, is enhanced by close association with the ribosomal proteins. Some of the ribosomal proteins bind directly to the rRNAs during transcription. The 28S, 18S, and 5.8S rRNAs are synthesized in the *nucleolar region* of the nucleus. The 5S rRNA is not transcribed in the nucleolus but rather from separate genes within the nucleoplasm. Processing of the rRNAs (see Section 18.5) includes cleavage to the functional size, limited formation of internal

Table 18.3 Characteristics of Regions in tRNA

Region	*Number of Nucleotides*	*Comments*
Amino acid helix	14 (7 base pairs)	Region where base mispairing occurs frequently; $G \cdot U$ is common; $-CCA_{OH}$ is added posttranscriptionally
Dihydrouracil stem	6 or 8 (3 or 4 base pairs)	First and last base pairs are usually $C \cdot G$
Dihydrouracil loop (loop I)	7–10	Region exhibits considerable variation
Anticodon stem	10 (5 base pairs)	Second base pair from anticodon loop is usually $C \cdot G$; from 5′ end, the 3rd, 4th, and 5th bases are the anticodon; 5′ side of anticodon is always a pyrimidine; 3′ side is usually a modified purine
Anticodon loop (loop II)	7	
Variable arm (loop III)	3–21	Extremely variable in structure and often lacks a helical stem; the arm probably forms hydrogen-bonded stem region (3–7 base pairs)
TΨC stem	10 (5 base pairs)	Base pair adjacent to TΨC loop is $C \cdot G$
TΨC loop (loop IV)	7	All tRNAs contain the sequence T-Ψ-C-purine at the same location in the loop; the purine is usually guanine

base pairing via hydrogen bonds, modification of particular nucleosides, and association with ribosomal proteins to form a stable tertiary conformation. Due to the complex nature of the ribosome, most studies on the rRNAs have been done in solution rather than in situ.

The 5S rRNA is 68% base-paired, with helical regions formed

Table 18.4 Ribosomal RNA Characteristics

Sedimentation Value	*Molecular Weight*	*G + C (%)*	*Methylations*	*Number of Gene Copies*
Eucaryotes				
p45S (nucleolar)	4.3×10^6	70	Mostly 2′-*O*-methylribose	500–1,000
			All retained in	
p41S	3.1×10^6	~70	28S and 18S	—
p32S	2.1×10^6	70		—
p20S	0.95×10^6	<70		—
m28S	1.7×10^6	65	9–13/1,000 NT	—
m18S	0.7×10^6	58	15–19/1,000 NT + m^6_2A	—
m5.8S	5.1×10^4	~65	?	—
m5S (nuclear)	3.9×10^4	65	None	Several hundred
Mitochondrial				
16S	5.4×10^5		Very low	1
12S	3.5×10^5		Very low	1
Procaryotes				
23S	1.1×10^6		11/1,000 NT	5–6
16S	0.6×10^6		17/1,000 NT	5–6
5S	4×10^4	65	None	~7

(p, precursor form of the RNA; m, mature form of the RNA)

between proximal as well as distal internal complementary sequences (see Figure 18.5). The nucleotide sequences and proposed conformations for 5S rRNA have been highly conserved throughout the evolutionary scale. The length, some sequences, and most helical regions are the same for *E. coli* and human 5S rRNA. One sequence is complementary to the —TΨC— loop in tRNAs; however, there is no direct evidence for a tRNA–5S rRNA interaction. In ribosome assembly the 5S rRNA associates with the 28S rRNA and several ribosomal proteins to form preribosomal particles. A specific function for 5S rRNA has not been described, although it is apparently required, in a structural role, for protein synthesis. A lack of 5S rRNA in a ribosome or cleavage at specific locations in the 5S rRNA nucleotide chain render the ribosome inactive.

The 5.8S rRNA, with 158 nucleotides, is closely associated, by hydrogen bonds, with the 28S rRNA. The 5.8S rRNA has considerable internal base-pairing, while the 5′ and 3′ ends are free to interact with the 28S rRNA. Like the 5S rRNA, the 5.8S rRNA nucleotide sequence has been conserved during evolution, but the 5.8S rRNA is not found in procaryotes.

The larger rRNAs contain most of the altered nucleosides found in

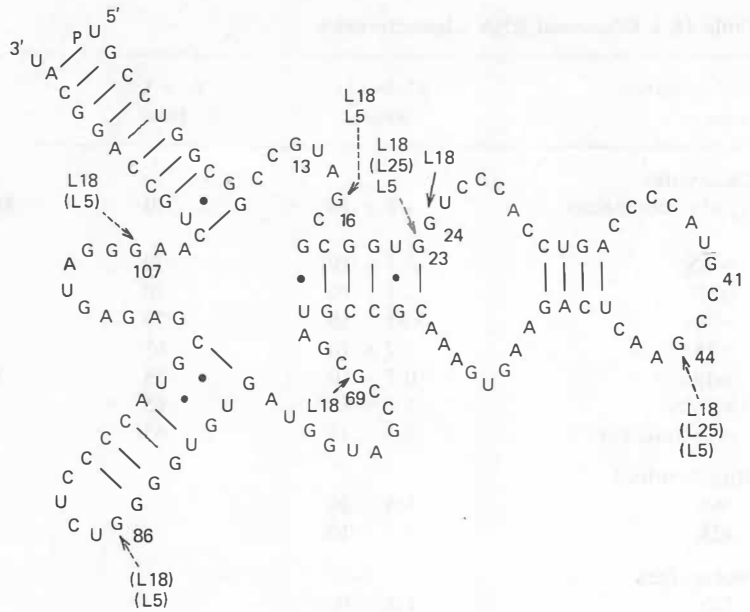


Figure 18.5
Secondary, base-paired, structure proposed for 5S rRNA.
Arrows indicate regions protected by proteins in the large ribosomal subunit.
Combined information from G. E. Fox and C. R. Woese, *Nature*, 256:505, 1975, and R. A. Garrett and P. N. Gray.

rRNA. These are primarily methylations on the 2′ position of the ribose, giving 2′-*O*-methylribose. There are a small number of N^6-dimethyladenines present in the 18S rRNA. The 28S rRNA has about 45 methyl groups and the 18S rRNA has 30 methyl groups, which may be involved in processing of the 45S precursor molecule (see Table 18.4).

In addition to the base-pairing within each 18S and 28S rRNA there is evidence for a base sequence in mRNAs, which can base-pair with the rRNA of the smaller subunit, forming a translation complex. The hinging mechanism between the two ribosomal subunits, which enables translocation and mRNA movement, is thought to involve protein–protein interactions and base-pairing between the 18S and 28S rRNAs.

Messenger RNA

The mRNAs are the direct carriers of genetic information from the nucleus to the cytoplasmic ribosomes. Each eucaryotic mRNA con-

tains information for only one polypeptide chain, and therefore these mRNAs have been designated monocistronic, whereas in prokaryotes mRNA can exist as polycistronic molecules. A cell's phenotype and functional state is related directly to the cytoplasmic mRNA content. In cells exhibiting highly active protein synthesis, such as pancreatic cells, the DNA : RNA ratio is very low due to the large amounts of mRNA and rRNA. However, in cells with low rates of protein synthesis, such as muscle cells, the DNA : RNA ratio is much higher, since there is no requirement for large quantities of mRNAs and ribosomes.

In the cytoplasm mRNAs have relatively short life spans, which in part are determined by the cell's particular needs at any given time. The life spans of mRNAs vary from only a few hours to several days or weeks. For example, the globin mRNAs for hemoglobin have to be very stable, since human erythrocytes continue to synthesize hemoglobin for weeks following loss of the cellular nucleus. In procaryotes mRNA half-life can be as short as 1 min. Some mRNAs are known to be synthesized and stored in an inactive or dormant state in the cytoplasm, ready for a quick protein synthetic response. An example of this is the unfertilized egg of the African clawed toad, *Xenopus laevis*. Immediately upon fertilization the egg undergoes rapid protein synthesis, indicating the presence of preformed mRNA and ribosomes.

Eucaryotic mRNAs have unique structural features not found in rRNA or tRNA (see Figure 18.6). These features aid in proper mRNA functioning. Since the information within mRNA lies in the

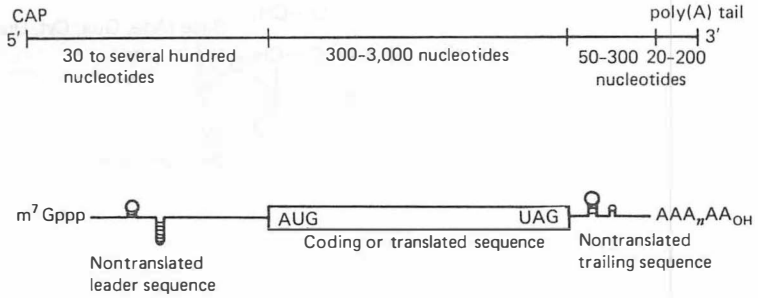


Figure 18.6
General structure for a eucaryotic mRNA.
There is a "blocked" 5′ terminus, cap, followed by the nontranslated leader containing a promoter sequence. The coding region usually begins with the initiator codon AUG and continues to the translation termination sequence UAG, UAA, or UGA. This is followed by the nontranslated trailer and a poly(A) tail on the 3′ end.

linear sequence of the nucleotides, the integrity of this sequence is extremely important. Any loss or change of nucleotides could alter the protein being translated. The translation of mRNA on the ribosomes must also begin and end at specific sequences. Structurally, starting from the 5′ terminus, there is an inverted methylated base attached via 5′-phosphate-5′-phosphate bonds rather than the usual internucleotide 3′,5′ phosphodiester linkages between adjacent riboses. This structure, called a "cap," is a guanosine 5′-triphosphate methylated at the number 7 nitrogen ($m^7G^5'ppp$). The cap is attached to the first transcribed nucleotide, usually a purine, methylated on the 2′-OH of the ribose (see Figure 18.7). The cap is followed by a nontranslated or "leader" sequence to the 5′ side of the coding region. The leader region is of variable length and includes the ribosome binding or attachment site. Following the leader sequence is the initiation sequence or codon, most often

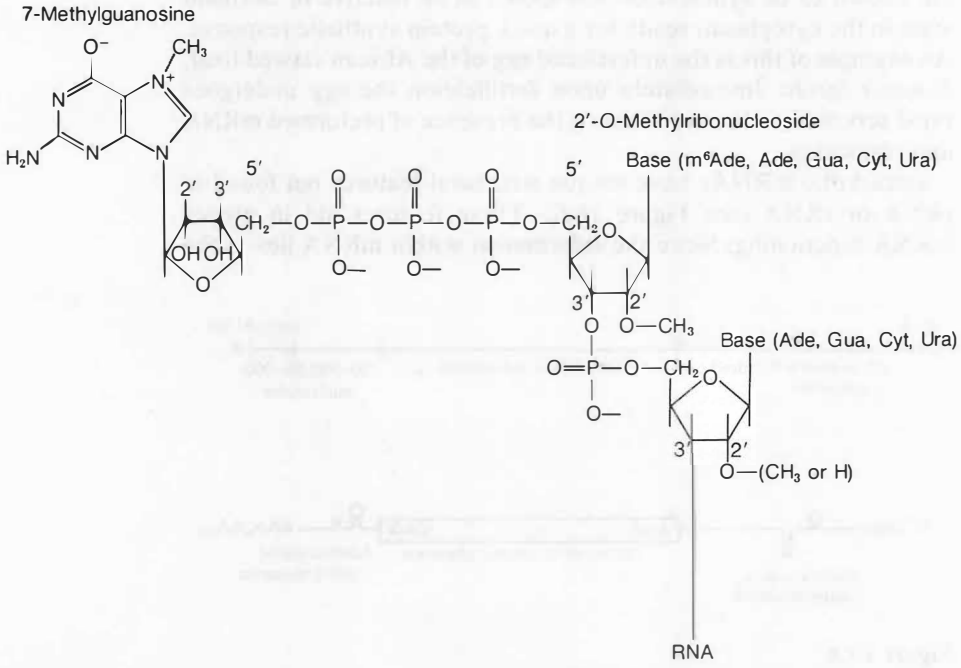


Figure 18.7
Diagram of the "cap" structure or blocked 5′ terminus in mRNA.
The 7-methylguanosine is inverted to form a 5′-phosphate to 5′-phosphate linkage with the first nucleoside of the mRNA. This nucleoside is often a methylated purine.

AUG, and then the translatable message or coding region of the molecule. At the end of the coding sequence is a termination sequence signaling nascent peptide and ribosome release. A second nontranslated or "trailer" sequence follows, terminated by a string of adenylic acids, called a poly(A) tail, which makes up the 3′ terminus of the mRNA. This poly(A) section may vary from 20 to 200 nucleotides.

The 5′ cap structure blocks the action of RNA exonucleases and phosphatases, which could attack the 5′ terminus of the message. The cap also has a positive effect on the initiation of message translation. The m^7GpppN structure has a rigid conformation, enhancing base stacking, which promotes the formation of stable initiation complexes. A message with a 5′-terminal m^7GpppN is translated more efficiently than one with a nonmethylated 5′-GpppN or a noncapped 5′-pppN. Several methylated nucleosides occur in the internal portions of some mRNAs in addition to those on the cap and adjacent nucleotides. The majority of the internal methylations are m^6-adenosines with some m^5-cytidines. Hemoglobin mRNA does not have internal methylations.

The role for the poly(A) tail is unclear. Several possibilities exist and the actual function may include some, all, or none of these: the poly(A) may function in transport of mRNA through the nuclear membrane; it may serve as a buffer for exonucleolytic attack from the 3′ terminus; or it may have some role in translation by interacting with the ribosome or a nonribosomal protein. Although the majority of mRNAs have a poly(A) tail, some may not. For example, the messengers for some histones lack 3′-poly(A) termini.

There are two types of protein associated closely with the eucaryotic mRNAs. One type binds tightly to the 5′ cap region and may be cap specific. The molecular weights of the polypeptides vary in each species studied. One of these 5′-RNA binding proteins has been identified as the translation factor eIF-4b. The second type of protein is associated with the 3′ end of the mRNAs.

Nuclear RNA

The nucleus, at some time or another, contains all of the cell's RNA, since all RNA is transcribed from a DNA template. These primary transcription products may be divided into two general classes: those that are precursor molecules for the cytoplasmic mRNAs, tRNAs, and rRNAs, and those RNAs that remain and function within the nucleus.

Examination of total nuclear RNA reveals a group of extremely large RNAs. This group has a rapid turnover rate, a size range from 30S to 100S, and has been called heterogeneous nuclear RNA,

HnRNA (see Table 18.1). There is sufficient evidence to identify many of these as precursors to mRNAs. Those HnRNAs destined to become mRNAs have 5′ caps, 3′-poly(A) sequences, and have been shown to hybridize with complementary DNA (cDNA), made from cytoplasmic mRNA, using reverse transcriptase. Mechanisms proposed for processing these large HnRNAs are described in Section 18.5 on posttranscriptional processing. It is generally thought that only one mRNA is contained within each HnRNA. Like other RNAs, HnRNA is complexed with specific proteins to form nuclear ribonucleoprotein particles (nuclear mRNP). The proteins within these nuclear mRNP are not ribosomal or chromosomal proteins.

Another nuclear RNA, the 45S precursor to the 18S, 28S, and 5.8S rRNAs, is synthesized in the nucleolus and processed in the nucleoplasm. In addition to these larger precursors there are the precursors for cytoplasmic 5S rRNA and tRNAs.

Some RNAs found in the nucleus and nucleolus tend to remain there and are relatively stable. These are the several small nuclear RNAs (*snRNA*), which range in size from 100 to 300 nucleotides and are not involved directly with protein synthesis. Some of the snRNAs are associated with chromatin and are thought to have structural and regulatory roles. Other snRNAs form ribonucleoprotein particles with specific nuclear polypeptides, which may be involved in posttranscriptional RNA splicing. These particles have been implicated in systemic lupus erythematosus.

Mitochondrial RNA

Mitochondria have their own protein-synthesizing mechanisms, including ribosomes, tRNAs, and mRNAs. The mt rRNAs, 12S and 16S, are transcribed from the mitochondrial DNA (mtDNA) as are at least 19 specific tRNAs and some mRNAs. The mt RNAs account for 4% of the total cellular RNA. They are transcribed by a mitochondrial-specific RNA polymerase. mt tRNAs are not as highly methylated as their cytoplasmic counterparts and are processed from precursor molecules smaller than their cytoplasmic counterparts. Genes for 12 tRNAs are located on the heavy mtDNA strand and 7 on the light strand. Some of the mRNAs have eucaryotic characteristics, such as 3′-poly(A) tails. A large degree of coordination exists between the nuclear and mitochondrial genomes. Most of the aminoacylating enzymes for the mt tRNAs and most of the mitochondrial ribosomal proteins are specified by nuclear genes, translated in the cytoplasm and transported into the mitochondria.

18.4 MECHANISMS OF TRANSCRIPTION

Transcription is the synthetic process that transfers genetic information from DNA to RNA molecules. Transcription occurs in the nucleoplasm, the nucleolus, and the cytoplasmic organelles containing DNA. The enzymes responsible for this irreversible process are called DNA-dependent RNA polymerases.

Molecular Components Required for Transcription

Template

Transcription occurs from a *DNA template* and for any given gene sequence, only one strand of the DNA is transcribed. Therefore, transcription is considered to be asymmetric.

Enzyme

DNA-dependent RNA polymerase catalyzes the polymerization of the activated precursors. This enzyme uses all four nucleoside triphosphates and is involved from initiation through termination of the process. RNA polymerase does not require a primer to initiate RNA synthesis as with DNA polymerase. Eucaryotes have several types of RNA polymerases due to the various DNA locations and the strict constraints on gene expression. Procaryotes and mitochondria possess only one form of RNA polymerase.

Precursors

The activated precursors are the ribonucleoside 5′-triphosphates, *ATP, GTP, CTP,* and *UTP.* All modifications of the bases and sugars occur after their polymerization into oligonucleotides.

Cofactors

RNA polymerases require Mg^{2+} and Mn^{2+}. In addition, *ATP* is used as an energy source. The specificity of transcription requires proper initiation sequences and termination sequences. A particular DNA sequence or template must be available to RNA polymerase at the desired time for gene expression to occur. Several proteins and structural RNAs involved in these regulatory steps are discussed in Chapter 20.

Basic Steps of Transcription

The process of transcription can be divided into four basic steps: *recognition, initiation, elongation,* and *termination.* The DNA-dependent RNA polymerase of the procaryote *E. coli* has been characterized extensively and used as a model for studying RNA polymerase in other procaryotes and eucaryotes. *E. coli* RNA polymerase exists as a *holoenzyme* containing five protein subunits. These subunits, designated α, β, β' and σ, form the protein complex $\alpha_2\beta\beta'\sigma$, having a molecular weight of about 476,000 (see Table 18.5). The holoenzyme promotes and initiates specific asymmetric transcription of a DNA template. The σ subunit only binds to DNA as part of the RNA polymerase. Following recognition of a correct promoter transcription site and initiation of RNA synthesis, the sigma subunit is released and the *core enzyme,* $\alpha_2\beta\beta'$, continues RNA elongation. The β' subunit is directly involved in maintaining the DNA : RNA polymerase complex. The early steps of transcription are diagrammed in Figure 18.8.

Recognition

RNA polymerase recognizes and attaches to discrete sites on the DNA genome. The accessibility of these binding sites, called promoters, depends upon the physiologic state of the cell. The binding sites are often the target of various regulatory factors and RNAs, which may act as positive or negative effectors. A positive effector enhances the binding of RNA polymerase and promotes transcription. The exact nature of the binding sites and the interactions with the effector molecules are not clear. In procaryotes there is a small peptide, the *sigma* subunit, which interacts with the core RNA polymerase, reduces nonspecific binding of the polymerase to DNA, and enhances the specific binding. This enhancement involves a change in the configuration of RNA polymerase and DNA, with the DNA changing to an "open" structure. The initial polymerase binding occurs through electrostatic forces. The specific "open" complex has a strong association constant, and a bound half-life of several hours. The nonspecific or "closed" configuration binding is less stable.

Since chromatin is a highly coiled and protected structure, specific interactions with RNA polymerase require relaxation of the DNA superhelix and localized denaturation of the hydrogen bonds between the DNA strands associated with the RNA polymerase. When DNA associated with RNA polymerase is treated with DNA nucleases, it becomes possible to obtain some knowledge aboutthe size and structure of the RNA polymerase recognition site. Approximately 40 base pairs are protected from the action of nuclease but

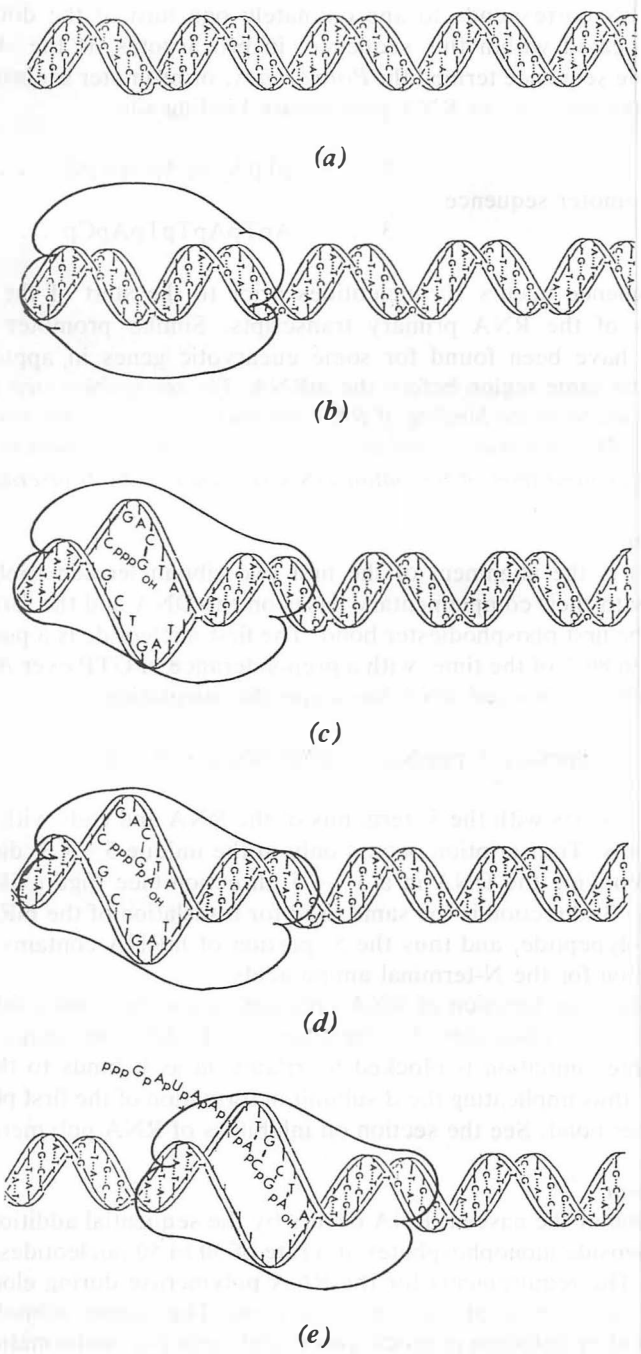


Figure 18.8
Early events in transcription.
(a) DNA double helix. (b) Recognition: RNA polymerase with "sigma" factor binds to a DNA promoter region in a "closed" configuration. (c) Initiation: The complex is converted to an "open" configuration, and the first nucleoside triphosphate aligns with the DNA. (d) The first phosphodiester bond is formed and the "sigma" factor released. (e) Elongation: Synthesis of nascent RNA proceeds with movement of the RNA polymerase along the DNA. The double helix reforms.

only 6 to 11 of these are "open" or denatured single-stranded regions. This corresponds to approximately one turn of the double helix. Located within this sequence, in procaryotes, is the short nucleotide sequence termed the *Pribnow box*, or promoter sequence, which may serve as an RNA polymerase binding site.

5′ . . . pTpApTpApApTpG . . . 3′

DNA promoter sequence

3′ . . . ApTpApTpTpApCp . . . 5′

This sequence occurs 10 nucleotides prior to the start of the 5′-terminus of the RNA primary transcripts. Similar promoter sequences have been found for some eucaryotic genes in approximately the same region before the mRNA. *The recognition step can be summarized as the binding of RNA polymerase to a specific site on the DNA. This step is mediated by a protein cofactor and results in the localized denaturation of the initial DNA sequence to be transcribed.*

Initiation

Initiation is the alignment of the first two ribonucleoside triphosphates with their complementary bases on the DNA and the formation of the first phosphodiester bond. The first nucleotide is a purine more than 80% of the time, with a preponderance of GTP over ATP. The newly synthesized RNA has a specific orientation:

$$pppG_{OH} + pppN_{OH} \rightarrow pppGpN_{OH} + P \sim P$$

Synthesis starts with the 5′ terminus of the RNA and ends with the 3′ terminus. Transcription occurs only in the unique 5′ → 3′ direction by copying the DNA in a 3′ → 5′ direction (see Figure 18.9). The 5′ → 3′ direction is the same used for translation of the mRNA into a polypeptide, and thus the 5′ portion of mRNA contains the information for the N-terminal amino acids.

The initiation function of RNA polymerase can be readily differentiated from elongation by treatment with different drugs. In procaryotes initiation is blocked by rifamycin as it binds to the β subunit, thus implicating the β subunit in formation of the first phosphodiester bond. See the section on inhibitors of RNA polymerase.

Elongation

Elongation of the nascent RNA occurs by the sequential addition of ribonucleoside monophosphates at a rate of 30 to 50 nucleotides per second. The requirements for the RNA polymerase during elongation are less stringent than for initiation. The sigma subunit is released after initiation in procaryotes, and there is a conformational

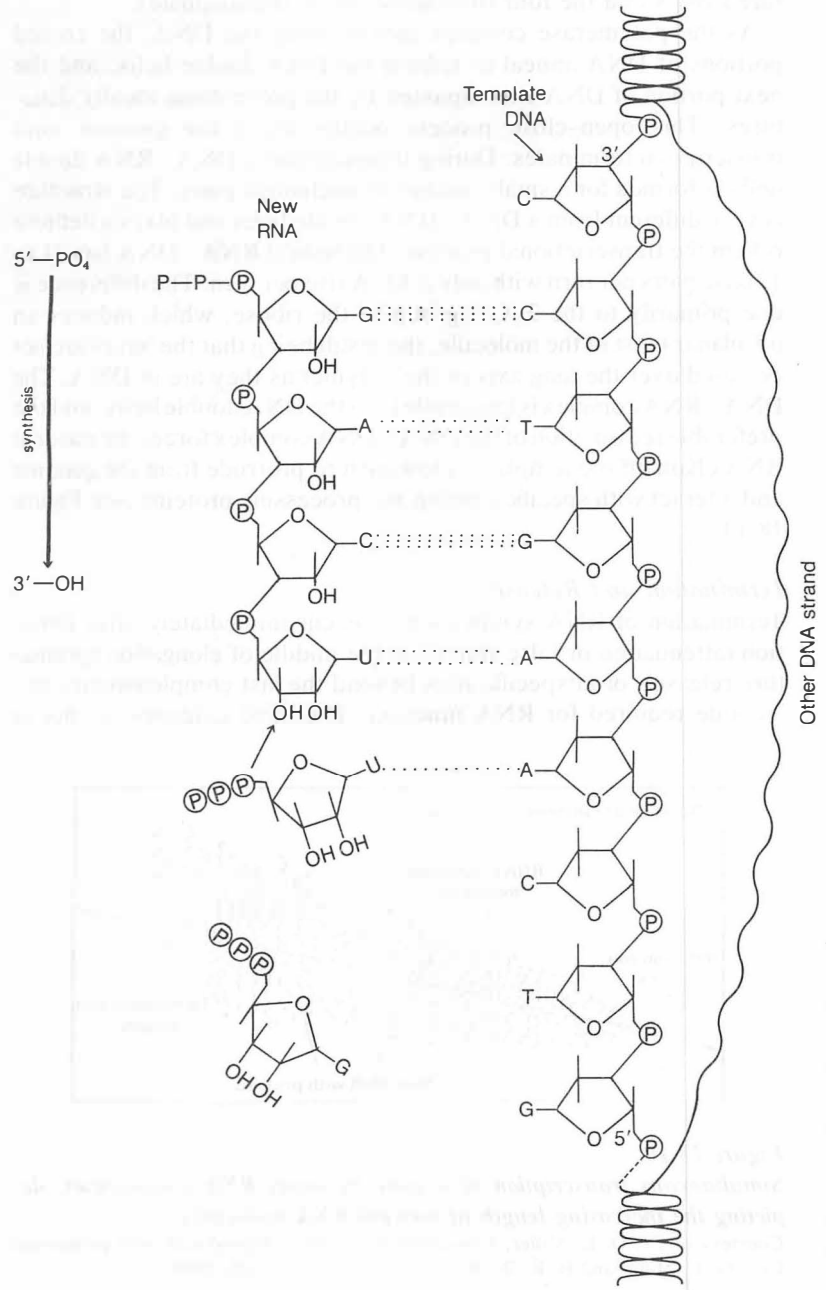


Figure 18.9
Biosynthesis of RNA showing asymmetry in transcription.
Nucleoside 5′-triphosphates align with complementary bases on one DNA strand, the template. RNA polymerase catalyzes the formation of the 3′,5′ phosphodiester links by attaching the 5′-phosphate of the incoming nucleotide to the 3′-OH of the growing nascent RNA releasing P_i. The new RNA is synthesized from its 5′ end toward the 3′ end.

change in the RNA polymerase. For elongation to continue, the transcription complex requires only complementary bases on the DNA, translocation of the RNA polymerase along the locally denatured DNA and the four ribonucleoside-5′-triphosphates.

As the polymerase complex moves along the DNA, the copied portions of DNA anneal to reform the DNA double helix, and the next portion of DNA encompassed by the polymerase locally denatures. This open–close process occurs along the genome until transcription terminates. During transcription a DNA · RNA double helix is formed for a small number of nucleotide pairs. The structure is very different from a DNA · DNA double helix and plays a definite role in the transcriptional process. The hybrid RNA · DNA has 11 to 12 base pairs per turn with only a 3.0 Å rise per turn. The difference is due primarily to the 2′-OH group of the ribose, which induces an off-planar twist of the molecule, the result being that the bases are not centered over the long axis of the polymer as they are in DNA. The DNA · RNA complex is less stable than the DNA double helix, and the preferable reformation of the DNA · DNA complex forces the nascent RNA chain off the template, allowing it to protrude from the genome and interact with specific binding and processing proteins (see Figure 18.10).

Termination and Release

Termination of RNA synthesis may occur immediately after initiation (attenuation or false starts), in the middle of elongation (premature release), or at specific sites beyond the last complementary nucleotide required for RNA function. The third category results in

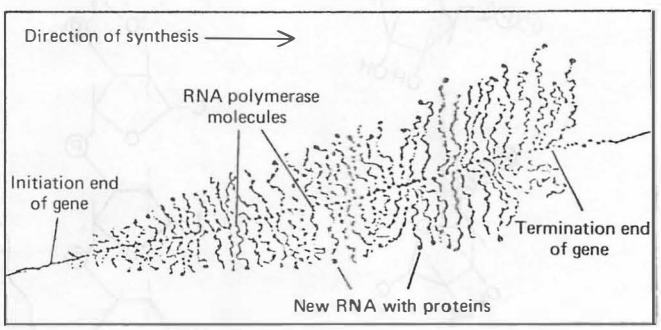


Figure 18.10
Simultaneous transcription of a gene by many RNA polymerases, depicting the increasing length of nascent RNA molecules.
Courtesy of Dr. O. L. Miller, University of Virginia. Reproduced with permission from O. L. Miller and B. R. Beatty, *J. Cell Physiol.*, 74:225, 1969.

correct RNAs of discrete size and information content. While termination sites have general sequence similarities, nothing universally specific has been delineated. In the RNA there usually are stretches of 4 to 8 uracils preceded by 4 to 12 guanines and cytosines. Since DNA GC sequences are transcribed more slowly than the DNA AT sequences, these GC stretches may serve to cause a "pause" in elongation, while the uracil stretch in the new RNA facilitates denaturation of the nascent RNA from the RNA polymerase DNA–RNA complex. Each of the known procaryote sequences can form a hairpin loop, which contains many G · C base pairs. Termination proteins that bind to the DNA or RNA polymerases and induce RNA release have been isolated from procaryotic systems. One of these, the *rho* factor, has a mass of 50,000 daltons, possesses ATPase activity, and forms a *rho*–DNA complex, which interacts directly with RNA polymerase. Similar proteins have been isolated from eucaryotic cells.

Asymmetry of Transcription

In a previous section it was stated that DNA stores genetic information in only one of the strands of the duplex DNA. Therefore RNA should be transcribed from only one strand of the DNA. This apparently is the situation, although there do not appear to be any restrictions on which strand contains the information. For example, one gene or transcriptional unit may have information in one strand, while an adjacent gene will have the information in the opposite strand. This situation can be considered to be constantly in flux, since the frequency of crossing-over or gene recombination is relatively high in eucaryotes. Selective transcription conserves energy as well as substrate.

Mitochondrial transcription is an exception to the concept of single-strand copying. In the small, circular mtDNA, RNA is copied from both DNA strands. The functional RNA molecules are processed from the larger primary transcripts with the remainder being degraded. The overall result is that the information retained is copied from only one DNA strand, as it is in the nucleus.

Proof for a single-stranded DNA template has been obtained for many RNA transcripts, using a variety of techniques. Early experiments designed to prove that RNA was copied from DNA measured the molar ratios of the nucleotides and analyzed nearest-neighbor frequencies. Similarities between the RNA and DNA were sufficient to implicate DNA as the source of RNA sequences. Recent advances in RNA, DNA, and protein sequencing have made the evidence for asymmetrical transcription more firm. It is now possible to compare amino acid sequences of polypeptides directly with nucleotide sequences and show whether the two are compatible. The

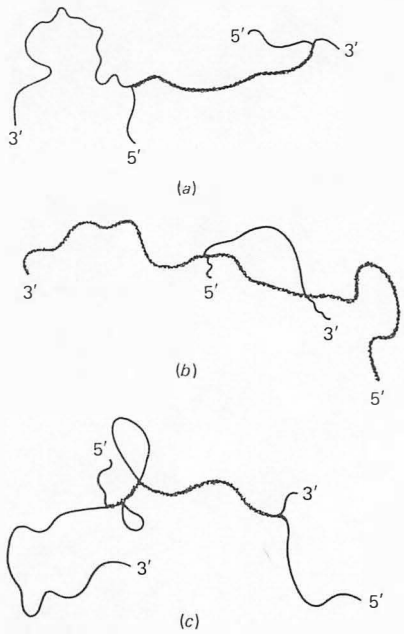


Figure 18.11
Examples of DNA · DNA and DNA · RNA hybridization products as shown by tracings of electron microscope photographs.
(a) DNA · DNA or DNA · RNA hybridization product formed between a long piece of single-stranded DNA and a shorter complementary DNA or RNA fragment. (b) Hybridization of double-stranded DNA with a smaller piece of cDNA, indicating complementarity with only one DNA strand (see text for further explanation). (c) Hybridization product formed from a long piece of single-stranded DNA annealed with a complementary strand of spliced RNA. In this case the RNA is now shorter than the transcriptional unit and primary transcript, lacking at least two internal sequences, as evidenced by the two loops of unpaired DNA.
Reprinted with permisson from H. Westphal, *Miami Winter Symp.*, 16:253, 1979.

best evidence, however, comes from hybridization of separated and isolated DNA strands with RNA or a complementary DNA copy of the RNA (cDNA). If the RNA or cDNA sequence is complementary to a piece of DNA, a double-stranded structure can be seen with the electron microscope (Figure 18.11*a*). If denatured, double-stranded DNA is hybridized with cDNA, then a D loop can be detected, in which case the cDNA probe is base-paired to one of the DNA strands, leaving the other DNA strand unpaired and looped out from the double helix (see Figure 18.11*b*). Likewise, if the DNA sequence includes more nucleotides than the RNA, a loop of single-strand DNA will be forced out of the linear hybridized complex (Figure 18.11*c*).

DNA-Dependent RNA Polymerase

Cellular RNA synthesis can be catalyzed by DNA-dependent RNA polymerase, RNA-dependent RNA polymerase, and polynucleotide phosphorylase. Transcription occurs only through the action of the DNA-dependent RNA polymerase.

Eucaryotes have multiple DNA-dependent RNA polymerases, which function in close association with regulatory proteins. The RNA polymerases may be divided into three classes, according to structure, cellular location, and function. In addition, the RNA polymerases can be separated by chromatographic procedures and respond differently to the toxin, α-amanitin (Table 18.5).

RNA polymerase I, or A, is located in the nucleolus and transcribes the ribosomal RNAs 18S, 28S, and 5.8S. It is insensitive to α-amanitin, except at very high concentrations.

RNA polymerase II, or B, functions in the nucleoplasm and is responsible for the synthesis of the HnRNAs and precursor mRNAs. RNA polymerase II can also transcribe viral mRNAs. This enzyme is very sensitive to α-amanitin and can be inhibited by concentrations of the drug as low as 10^{-9} M. There are at least four forms of RNA polymerase II, with different subunit compositions, which are found in different types of tissues. RNA polymerase II responds quickly and accurately to regulation pressures, which signal changes in gene expression.

RNA polymerase III, or C, also exists in several forms, having different subunit compositions. These enzymes are located in the nucleoplasm and are responsible for the synthesis of 5S rRNA and the precursors to the tRNAs. Class III polymerases can be inhibited by moderate amounts of α-amanitin, 10^{-4} to 10^{-5} M.

All the eucaryotic RNA polymerases have complex structures which reflect their multitude of functions. They are all metalloproteins requiring the metal ions Mg^{2+} and Mn^{2+}. Unlike DNA

Table 18.5 Comparative Properties of Some RNA polymerases

	Nuclear				
	I (A)	*II (B)*	*III (C)*	*Mitochondrial*	E. coli
High mol wt	195–197	240–214	155	65	160 (β')
subunits[a]	117–126	140	138		150 (β)
Low mol wt	61–51	41–34	89		86 (σ)
subunits	49–44	29–25	70		40 (α)
	29–25	27–20	53		10 (ω)
	19–16.5	19.5	49		
		19	41		
		16.5	32		
			29		
			19		
Variable forms	2–3 types	3–4 types	2–4 types	1	1
Specialization	Nucleolar; rRNA	mRNA	tRNA	All mtRNA	None
		Viral RNA	5S rRNA		
Inhibition by α-amanitin	Insensitive (>1 mg/ml)	Very sensitive 10^{-9}–10^{-8} M	Sensitive 10^{-5}–10^{-4} M	Insensitive, but sensitive to rifampicin	Rifampicin-sensitive

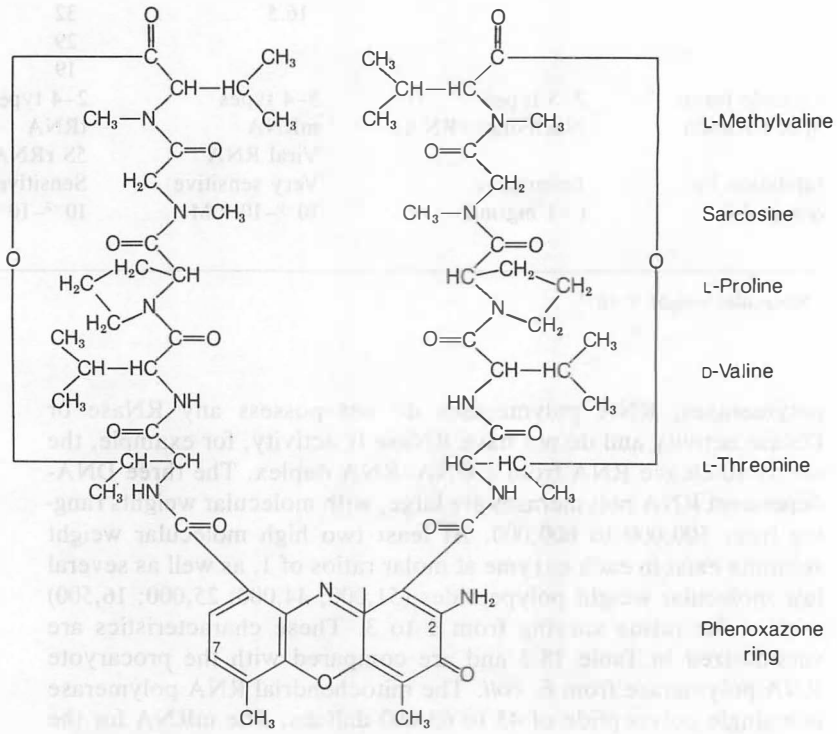
[a] Molecular weight × 10^{-3}.

polymerases, RNA polymerases do not possess any RNase or DNase activity and do not have RNase H activity, for example, the ability to cleave RNA from a DNA–RNA duplex. The three DNA-dependent RNA polymerases are large, with molecular weights ranging from 500,000 to 600,000. At least two high molecular weight subunits exist in each enzyme at molar ratios of 1, as well as several low molecular weight polypeptides (51,000; 44,000; 25,000; 16,500) with molar ratios varying from 1 to 3. These characteristics are summarized in Table 18.5 and are compared with the procaryote RNA polymerase from *E. coli*. The mitochondrial RNA polymerase is a single polypeptide of 45 to 65,000 daltons. The mRNA for the mitochondrial RNA polymerase is transcribed from a nuclear gene. Like procaryote RNA polymerases the action of this enzyme is inhibited by the rifamycin family of antibiotics, for example, rifampicin, and insensitive to α-amanitin.

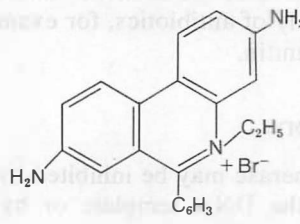
Inhibitors of Transcription

The functions of RNA polymerase may be inhibited by compounds that interact directly with the DNA template or by compounds that bind directly to the polymerase.

DNA-specific compounds bind directly to DNA and act by altering template availability. The common inhibitors in this category are planar molecules with multiple aromatic rings. They insert between base pairs (intercalate), span the double helix, and have a high affinity for intercalation between adjacent dG · dC pairs. Actinomycin D (Figure 18.12) is the best known of the DNA binding compounds. The phenoxazone ring intercalates and the cyclic peptide portions hydrogen-bond specifically with dG residues above and below the phenoxazone ring in the minor groove. Actinomycin D preferentially



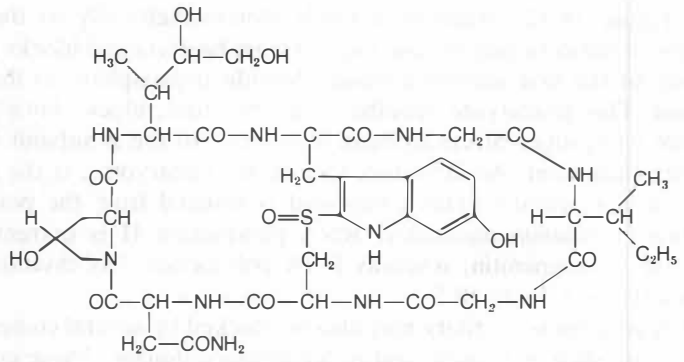
Actinomycin D



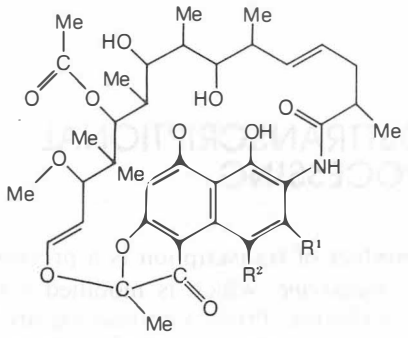
Ethidium bromide

Figure 18.12
Inhibitors of RNA synthesis.
Actinomycin D and ethidium bromide bind to DNA, blocking transcription. Rifamycin, rifampicin and α-amanitin bind directly to RNA polymerase.

Figure 18.12 (Continued)



α-Amanitin



Rifamycin B R^1 = H. R^2=O—CH_2—COOH

Rifampicin R^1 = —CH=N—N—CH_3. R^2=OH

blocks elongation and at different concentrations can have differential effects on RNA synthesis in the nucleus, nucleolus, and mitochondria (Clin. Corr. 18.1). Ethidium bromide (Figure 18.12) also inhibits elongation by intercalation. Since the ethidium bromide effect is concentration-dependent, it is possible to unwind and reverse the direction of a DNA circular supercoil if a high concentration of ethidium bromide is present in the solution. Small circular DNA is particularly sensitive to ethidium bromide, and therefore ethidium bromide can be used to inhibit mitochondrial transcription. Several other antibiotics that bind to DNA are chromomycin, mithramycin, daunomycin, and anthramycin.

There are only a few inhibitors of transcription known to act specifically on RNA polymerases. However, these inhibitors have been important for differentiating and classifying the RNA polymerases in prokaryotes and eukaryotes. The rifamycins (Clin. Corr. 18.1) have been used to study or inhibit procaryote transcrip-

CLIN. CORR. 18.1
ANTIBIOTICS

Rifamycin (rifampicin) is toxic to acid-fast and gram-positive bacteria and has been used to treat tuberculosis.

Actinomycin D is an effective bacteriocidal agent; however, its toxicity prevents it from becoming widely used as an antitumor compound, although it was used successfully for Wilm's tumor.

tion (Figure 18.12). Rifampicin binds stoichiometrically to the β subunit of the core polymerase ($\alpha_2\beta\beta'$) from bacteria and blocks the binding of the first incoming ribonucleoside triphosphate to the β subunit. This procaryote specific drug, therefore, blocks initiation but not elongation. Streptolydigan also binds to the β subunit but inhibits elongation. An inhibitor, specific for eucaryotes, is the potent toxin α-amanitin. This compound is isolated from the poison mushroom *Amanita phalloides*. RNA polymerase II is extremely sensitive to α-amanitin, whereas RNA polymerase I is essentially insensitive (see Table 18.5).

RNA polymerase activity may also be blocked by several complex polyanions such as heparin and polyethylenesulfonate. These compete with DNA for binding the polymerases.

18.5 POSTTRANSCRIPTIONAL PROCESSING

The immediate product of transcription is a precursor RNA molecule, the *primary transcript,* which is modified subsequently to a mature functional molecule. Primary transcripts are longer than the final RNA, have no modifications on the bases or sugars, and thus contain only A, G, C and U residues. The primary transcript is copied from a linear segment of DNA, a *transcriptional unit,* between specific initiation and termination sites. Transcriptional units may contain information in at least three forms: (1) information that is contiguous or without interruption; (2) information that is discontinuous, having sequences coding for a single protein or RNAs that are interrupted by unwanted nucleotide stretches; and (3) information that is in a tandem or repeated form in which information for multiple molecules is linked together and at some later stage requires separation. A gene, therefore, may not necessarily be colinear with the nucleotide or amino acid sequence of the final gene product (see Figure 18.13).

The summation of all the enzymatic reactions leading to mature functional RNA molecules from primary transcripts is called RNA processing. Processing involves a variety of events, which include base modifications, sugar modifications, pyrimidine ring rearrangements, formation of helices and tertiary configurations, additions to the 5′ terminus, additions to the 3′ terminus, specific exonucleolytic cleavages, specific endonucleolytic cleavages, complex cleavages with splicing of pieces, and formation of RNA–protein complexes.

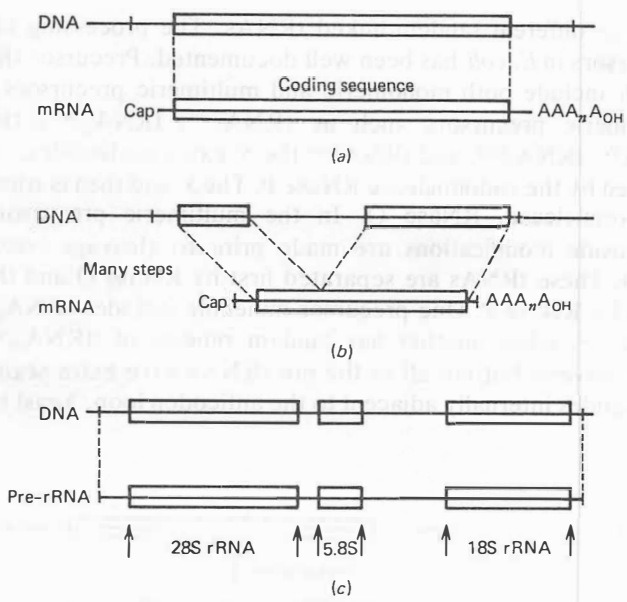


Figure 18.13
Models for transcription units in DNA showing coding regions and transcription products.
Arrows indicate cleavage points. (a) Linear information as expressed in histone genes. (b) Discontinuous information with subsequent splicing of intervening sequences as expressed in hemoglobin and ovalbumin genes. (c) Tandem information, single precursor with multiple mature RNAs: examples are rRNA and tRNA.

The number, type, and order of processing events is different for each group of RNA and often varies for each specific type within the groups.

Transfer RNA Processing

The proper recognition of mRNA by the tRNAs and the proper aminoacylation of the specific tRNAs are important and need to be rigorously controlled. Much of the specificity that results in overall tRNA function resides in the many nucleoside modifications (see Figure 18.4).

Cleavage

Primary transcripts for the tRNAs are trimmed to the proper size from both the 3′ and 5′ termini. These may be identical tRNA re-

peats or different tandem-linked tRNAs. The processing of tRNA precursors in *E. coli* has been well documented. Precursor tRNAs in *E. coli* include both monomeric and multimeric precursors. In the monomeric precursors such as $tRNA_I^{Leu}$, $tRNA_{III}^{Gly}$, $tRNA^{Asn}$, $tRNA^{Ile}$, $tRNA_{II}^{Glu}$, and $tRNA_I^{Asp}$ the 5′ extra nucleotides, ~15, are cleaved by the endonuclease RNase P. The 3′ end then is trimmed by an exonuclease, RNase Q. In the multimeric precursors most nucleoside modifications are made prior to cleavage (see Figure 18.14). These tRNAs are separated first by RNase O and then matured by RNase P. One precursor molecule includes $tRNA_{III}^{Ser}$ and $tRNA_{II}^{Arg}$, while another has tandem repeats of $(tRNA_{III}^{Gly})_n$. In yeast, several but not all of the pre-tRNAs have extra segments of nucleotides internally adjacent to the anticodon loop. Yeast $tRNA^{Tyr}$

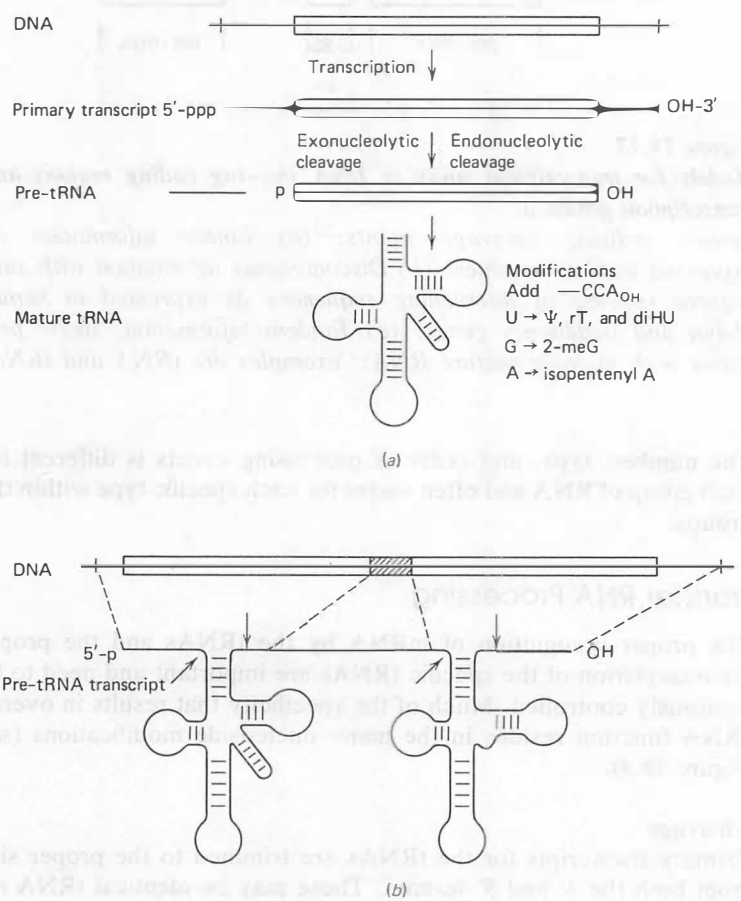


Figure 18.14
Processing of tRNA transcription products. *(a) Processing for single tRNA gene. (b) Processing for tandem tRNA genes. Arrows in (b) indicate cleavage points. Nucleoside modifications occur early in the maturation process.*

is transcribed from eight genes, and each transcript has 14 extra internal nucleotides that are removed. An extra internal sequence of 18 to 19 nucleotides is also transcribed and removed from yeast $tRNA^{Phe}$. Yeast cells have several clustered tRNA genes separated by only 9 nucleotides, which are cotranscribed. Similar structures occur in several eucaryotes, although extra internal (intervening) sequences for tRNAs have only been detected in yeast cells.

Additions

Each functional tRNA has at its 3′ terminus the sequence $—pCpCpA_{OH}$. In most instances this sequence is added sequentially by nucleotidyltransferase. Cells grown in the presence of actinomycin D, an antibiotic that blocks transcription, still add —CCA quickly to presynthesized tRNAs. Nucleotidyltransferase prefers ATP and CTP as substrates and always incorporates them into tRNA at a ratio of 2C/1A. $—CCA_{OH}$ ends are found on both cytoplasmic and mitochondrial tRNAs.

Modified Nucleosides

More than 60 different enzymic modifications to sugars and bases have been found in tRNAs (see Table 18.2). These occur at a high frequency in all tRNAs. In rat liver $tRNA^{Ser}$ there are 17 modified nucleosides of the 85 total nucleosides and 13 of the 17 are different. Many of these alterations are methylations and alkylations. Dihydrouridine is formed by reduction of the 5,6 carbon–carbon double bond in the uracil ring. Dihydrouridines are concentrated in the D loop at highly specific locations. Ribothymidine, one per tRNA, and pseudouridine, one or two per tRNA, are localized in the constant sequence —GTΨC— in the loop proximal to the 3′ terminus. Pseudouridine is unique from the other ribonucleosides because the linkage between the pyrimidine ring and ribose does not involve a ring nitrogen but rather the C-5 position of uracil. Formation of pseudouridine requires enzymatic cleavage of the β-glycosidic bond, rotation of the pyrimidine ring and reformation of the glycosidic link.

After the tRNA primary transcripts have been trimmed, spliced, methylated, folded, terminated, and so on, they are ready for their ultimate modification, an amino acid covalently attached to the 3′-terminal adenosine (see Chapter 19).

Ribosomal RNA Processing

The primary transcript for the ribosomal RNAs (28S, 18S, and 5.8S) is a large 45S precursor in the nucleolus. This is an example of multiple RNA molecules being transcribed in tandem and the functional sections cut out and retained, while the unwanted sequences

are degraded. The processing of the 45S rRNA precursor includes modification of nucleosides, cleavage, and strand-folding or base-pairing. The overall scheme for rRNA processing is depicted in Figure 18.15. In eucaryotic cells only 50–60% of the 4.6×10^6-dalton 45S precursor is actually retained as ribosomal rRNA.

As the 45S rRNA is transcribed, it associates with some ribosomal proteins and with the methyltransferases, which add the required methyl groups to the bases. Thus 80% of the total methyl groups on the mature molecule are on the 2′-OH of ribose. Methylation of the riboses occurs at a different cellular site. All methylations occur on the 45S rRNA, except for the formation of N^6-dimethyladenine on the 18S rRNA. The methyl groups are retained in the mature 28S and 18S rRNAs. The methylations, which use *S*-adenosylmethionine as the donor, are sensitive to actinomycin D. Pseudouridine, ~1.2–1.8 mol %, is also present in the 45S rRNA precursor.

Cleavage of the 45S rRNA occurs by sequential and specific endonucleolytic attack, yielding discrete intermediate RNAs. RNase P and RNase III activities for this type of cleavage have been detected in human cells. Base compositions have been determined for the 45S, 32S, 28S, and 18S RNAs. Whereas the 45S RNA has a G + C content near 70%, the 28S and 18S rRNAs have G + C contents similar to the cellular average of 56–65%. This indicates preferential degradation of G—C-rich regions (78% G + C). The unwanted regions may be bounded by methylated sequences serving as cleavage signals (see Table 18.4). Further stabilization of the rRNAs occurs by formation of base-paired helical hairpin loops. Many of the base-paired regions are retained in 28S rRNA. The remainder are degraded, but first may serve as processing signals for cleavage or ribosomal protein binding.

The locations of the 28S and 18S rRNAs within the 45S precursor rRNA have been determined by nucleotide sequence analysis, oligonucleotide mapping, and hybridization competition experiments, using the intermediate RNAs, 45S RNA, and nucleolar DNA. The 28S rRNA originates from the 3′ end of the 45S molecule, while the 18S rRNA is processed from the middle of the 5′ half of the 45S precursor.

Processing of rRNA in procaryotes also involves cleavage of high molecular weight precursors to smaller precursor and mature molecules (see Figure 18.15). Some of the bases are modified by methylation on the ring nitrogens of the bases rather than the ribose and by the formation of pseudouridine (only 0.06–0.3 mol %). The *E. coli* genome has approximately seven rRNA transcriptional units dispersed throughout the DNA. Each contains at least one 16S, one 23S, and one 5S rRNA or tRNA sequence. Processing of the rRNA is coupled directly to transcription, so that cleavage of a large pre-

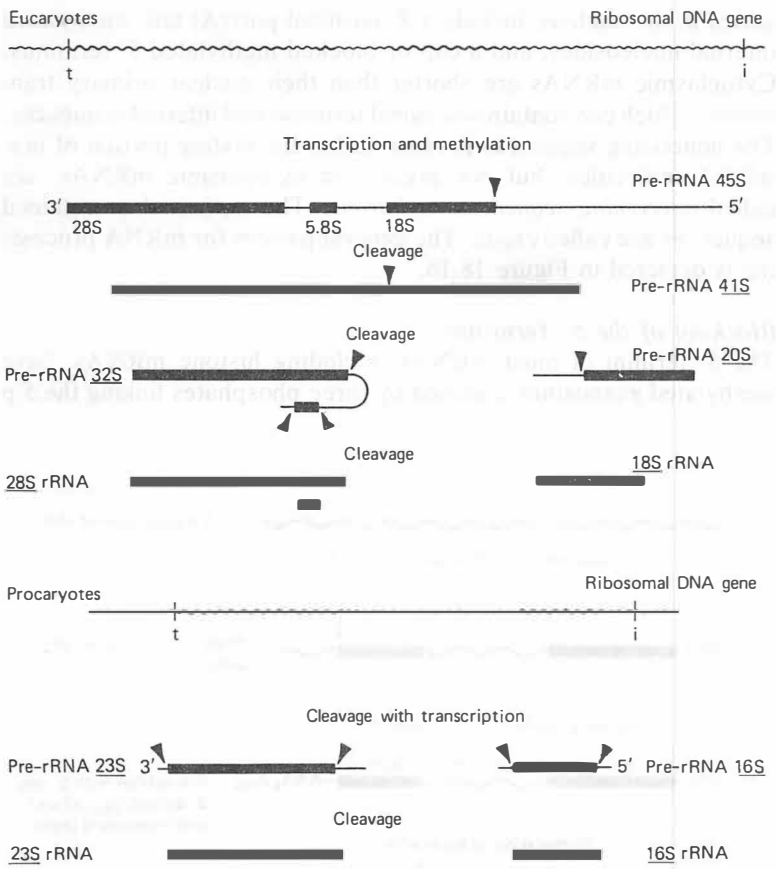


Figure 18.15
Schemes for transcription and processing of rRNAs.
Redrawn with permission from R. Perry, *Ann. Rev. Biochem.*, 45:611, 1976. Copyright 1976 by Annual Reviews, Inc.

cursor primary transcript, 30S, occurs at double-helical regions yielding precursor 16S, precursor 23S, precursor 5S, and precursor tRNAs. These precursors are slightly larger than the functional molecules and only require trimming for maturation.

Messenger RNA Processing

Most eucaryotic mRNAs have distinctive structural features, which are not a consequence of RNA polymerase action. These features,

added in the nucleus, include a 3′-terminal poly(A) tail, methylated internal nucleosides, and a cap or blocked methylated 5′ terminus. Cytoplasmic mRNAs are shorter than their nuclear primary transcripts, which can contain additional terminal and internal sequences. The noncoding sequences present within the coding portion of pre-mRNA molecules, but not present in cytoplasmic mRNAs, are called *intervening sequences* or *introns*. The *expressed* or retained sequences are called *exons*. The general pattern for mRNA processing is depicted in Figure 18.16.

Blocking of the 5′ Terminus

The 5′ termini of most mRNAs, including histone mRNAs, have methylated guanosines attached by three phosphates linking the 5′p

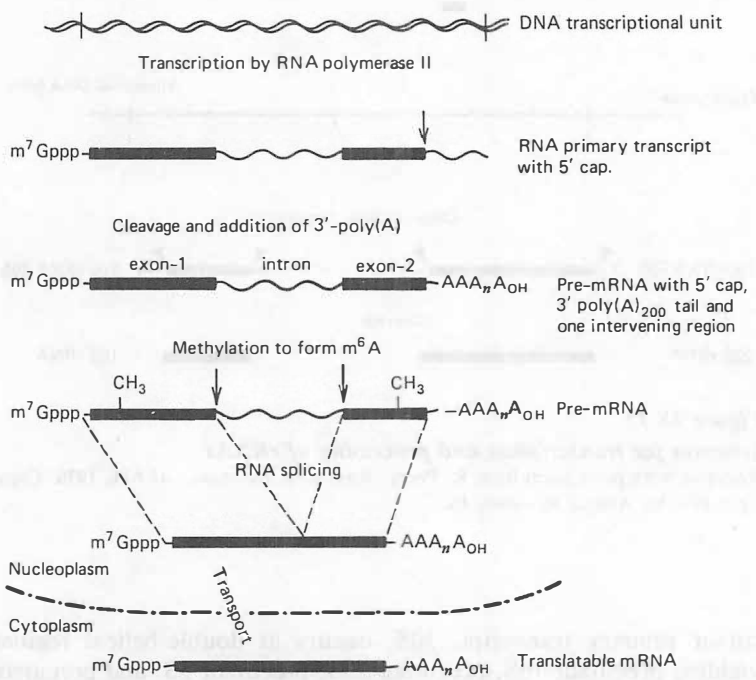


Figure 18.16
Scheme for processing mRNA.
The points for initiation and termination of transcription are indicated on the DNA. Arrows indicate cleavage points. The many proteins associated with the RNA and tertiary configurations are not shown.

of the cap nucleoside to the 5′p of the first transcribed nucleoside. The general form this takes is $m^7G^5ppp^{5'}N$. . . (see Figure 18.7). The cap is attached to a purine in 75% of the blocked mRNAs and to a pyrimidine in the other 25%.

Addition of the cap structures occurs on mRNAs or pre-mRNAs, and the enzymatic addition is associated with the initiation of transcription and RNA polymerase II (B). No capping is detected during transcription with RNA polymerases I (A) or III (C). The synthesis of the cap requires several allosteric enzyme activities. The steps in mRNA capping are outlined below.

1. Initiation of transcription

$$pppG + pppC \xrightarrow{\text{RNA polymerase II}} pppGpC + PP_i$$

2. Removal of inorganic phosphate to yield a 5′-di-phosphate

$$pppGpC \xrightarrow{\text{nucleotide phosphohydrolase}} ppGpC + P_i$$

3. Addition of the inverted or cap guanyl residue

$$pppG + ppGpC \xrightarrow{\text{guanyltransferase}} GpppGpC + PP_i$$

4. Methylation of the terminal guanine

$$GpppGpC + S\text{-adenosylmethionine} \xrightarrow{\text{methyltransferase}} m^7GpppGpC + S\text{-adenosylhomocysteine}$$

5. Methylation of the penultimate nucleoside

$$m^7GpppGpC + S\text{-adenosylmethionine} \xrightarrow{\text{methyltransferase}} m^7GpppGmpC + S\text{-adenosylhomocysteine}$$

The second methylation is usually on the 2′-OH position. Caps are conserved during the remainder of the processing and therefore found in cytoplasmic mRNAs.

Additions to the 3′ Terminus

A poly(A) tail is found on most cytoplasmic and many mitochondrial mRNAs. The larger nuclear RNAs, destined to become mRNAs, also have this modification. The string of adenylic acids (AMP) varies in length from 20 to 200 nucleotides. Poly(A) is synthesized in the nucleus by a poly(A) polymerase using ATP and the 3′-OH end of pre-mRNA as the substrates.

$$\text{RNA} + n\,\text{ATP} \xrightarrow[Mg^{2+}]{\text{poly(A) polymerase}} \text{RNA—(A)}_n + n\,\text{PP}_i$$

The addition of poly(A) can be inhibited by the nucleoside analog 3′-deoxyadenosine, cordycepin. This analog can be phosphorylated at the 5′ position and added by poly(A) polymerase to the end of mRNAs; however, the lack of a 3′-OH prevents further additions of AMP. Cells treated in this manner show a lack of newly synthesized mRNA transported to the cytoplasm. Separation of transcription events and poly(A) addition also has been demonstrated by treating cells with actinomycin D. Actinomycin D, at very low concentrations, blocks transcription but not the action of poly(A) polymerase, which adds AMP to the pre-mRNAs transcribed before addition of actinomycin D. Cytoplasmic addition of poly(A) to mRNAs is possible, but only in cells infected with certain DNA viruses, such as vesicular stomatitis virus (VSV) and vaccinia virus. In these situations the poly(A) polymerase is a viral enzyme. In mt mRNAs with 3′-poly(A) tails, the poly(A) is added by a mitochondrial specific poly(A) polymerase.

Cleavage and Splicing

The primary transcript for mRNAs may contain considerable noninformational nucleotide sequences. These sequences may appear in the cytoplasmic mRNAs immediately preceding the "start" codon for transcription and after the termination codon. The bulk of the extra nucleotides, however, are removed during processing. Extra nucleotides transcribed beyond the poly(A) site are trimmed by endonucleolytic cleavage and there may be a few nucleotides to the 5′ side of the cap sequence. The extra 5′ end trimming has been inferred for some mRNAs by the presence of a cap followed by a pyrimidine rather than a purine nucleotide.

Noninformational or intervening sequences are interspersed throughout the informational portion of the primary transcripts in many eucaryotes and hence were present in the DNA transcriptional unit for the gene. The presence of these sequences in DNA and primary transcripts and the loss of these sequences in cytoplasmic mRNAs may occur by two mechanisms.

1. *DNA recombination:* DNA may be rearranged prior to transcription so that widely separated DNA sequences become adjacent. An example of this may be the synthesis of mouse immunoglobulins and may explain the flexibility of the immune system. Typical DNA recombination mechanisms can be used to achieve this rearrangement (Figure 18.17). This type of control over mRNA sequences may be limited only to specialized systems, such as the

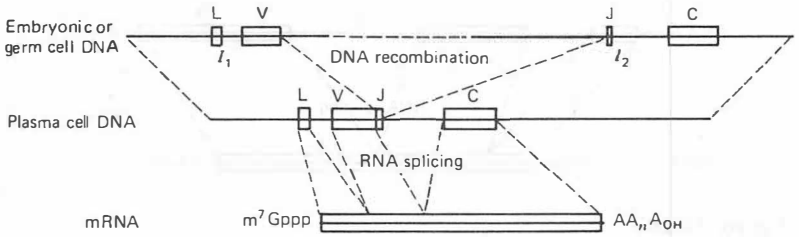


Figure 18.17

Diagram of a mechanism proposed for immunoglobulin gene rearrangements.

In embryonic or germ cells the coding sequences for the variable and constant regions of the peptide are located extremely far apart. In the differentiated cell, as indicated by nucleotide sequences of primary transcripts, these genetic regions are closer together. DNA recombination, in response to developmental pressures, may account for the genetic rearrangement. A primary transcript is made, which is subsequently spliced to form a mature mRNA. The leader sequence (L), the variable (V), and constant (C) coding regions, and the junction (J) sequence are indicated by open boxes. The regions between L and V and between J and C are intervening sequences found in both cell types and spliced out during RNA processing.

Courtesy of P. Leder.

synthesis of the immunoglobulins. The processing described below is probably more prominent.

2. *RNA splicing:* RNA polymerase transcribes a primary transcript containing exons (expressed sequences) and introns (intervening sequences). The introns are removed by endonucleolytic enzymes, and the exons are joined or spliced together forming a continuous informational sequence (see Figure 18.18). Splicing occurs in the nucleus probably after capping, methylation and poly(A) addition. Specific enzymes may be responsible for the recognition of intron–exon borders, cleavage, and ligation of the various pre-mRNAs. These mammalian enzymes have not been characterized, although one for yeast tRNA splicing has been isolated and requires Mg^{2+}, K^+, and ATP. The nuclease and ligase activities are independent. The yeast tRNA splicing enzyme does not seem to work on mRNAs.

The sequences of bases within the introns of a given primary transcript do not have any singular features and contain no apparent

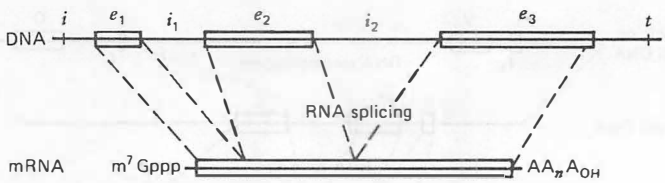


Figure 18.18
Diagram of RNA splicing showing the sequence of intervening and expressed regions in DNA.
A primary transcript is synthesized from i through t and processed to form an mRNA as indicated. The linearity of the exons is retained throughout. e_1, e_2, and e_3 refer to expressed regions, or exons. i and i_2 refer to intervening sequences, or introns.

genetic information. Intron lengths vary from 100 to 1,000 nucleotides for mRNAs and from 14 to 34 nucleotides for yeast tRNAs. The nucleotide sequences of the intron–exon borders have some similarities that appear to be common for several species. Introns in RNA generally begin with . . . pGpU. . . and end with . . . pApG. Those intron–exon borders sequenced have a tetranucleotide sequence (—pApGpGpU—) that allows for several splice points within the border while still retaining the correct mRNA sequence. Such a mechanism places less restriction upon the splicing enzyme and reduces mRNA loss during processing (see Figure 18.19). Procaryotes do not have the ability to process intervening sequences, and consequently most eucaryotic genes are not expressed in procaryotic hosts.

Examples of Eucaryote Spliced mRNAs
Mammalian hemoglobin mRNA was the first eucaryotic mRNA isolated. Whereas the globin mRNA is only 9S, the primary transcript for α-globin is 2.5 to 5 times longer than the mature mRNA. In the mouse the mRNA contains a cap, a 52-nucleotide leader sequence, the structural message for 144 amino acids, a 110-nucleotide nontranslated 3′ sequence, and a poly(A) tail. β-Globin pre-mRNAs with lengths of 5,000 and 1,500 nucleotides have been found. The half-life of the longer sequence is less than 10 minutes. Two introns have been located in the mouse, rabbit, and human β-globin primary transcripts. Although the intron sequences differ, their locations between codons 30-31 and codons 104-105 are identical for all three species. In mouse β-globin the first intron is about 646 nucleotides long and the second 116 nucleotides. The mouse α-globin gene also

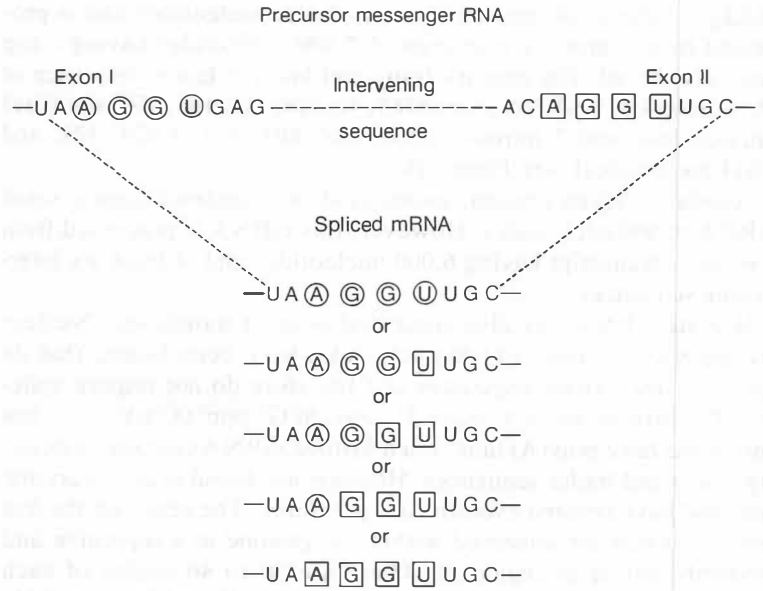


Figure 18.19
An example of an RNA splicing sequence and the way in which splicing can occur at several positions while yielding the same mRNA product.
Redrawn with permission from F. Crick, *Science*, 204:265, 1979. Copyright 1979 by the American Association for the Advancement of Science.

has two introns in positions similar to the β-globin genes (see Clin. Corr. 18.2 and 18.3).

Immunoglobulins are made up of four polypeptide chains, two light and two heavy. The light chain has variable regions and constant regions. Both DNA and RNA recombination are required for the expression of an immunoglobulin gene. The DNA genes are put in proximal positions by DNA recombination during differentiation and before transcription. However, the primary transcript still includes two introns, which are processed very efficiently. In mouse myeloma cells the globulin mRNAs are transcribed at ~20 molecules per minute, requiring 50 polymerases per gene. Each cell contains nearly 30,000 heavy chains and 40,000 light chains. Only 100 to 150 of these are found in the nucleus. The heavy chains are capped with $m^7Gppp(m^6)Ap^mAp^mC$. . . and the light chains with $m^7GpppGp^mAp^mA$ . . . . These mRNAs have a high metabolic stability with half-lives of 24 h.

Ovalbumin is synthesized in response to hormone stimulation in

CLIN. CORR. 18.2 ALTERED HUMAN GLOBIN GENES

Many clinical examples of altered globin gene expression exist in humans. These are due either to malfunctioning genes or incorrect RNA processing. Abnormal globin proteins, with a single amino acid change, may be produced, following point mutations in the coding region of the gene. Over 80 single amino acid substitutions have been identified. There may also be gene fusions as well as deletions of gene segments, which can cause other globin abnormalities. Hemoglobin Lepore, an extra-large globin molecule, results from the fusion of the δ- and β-globin genes. Another example of a fusion protein is hemoglobin Kenya. Both of these are translated from a large mRNA. Neither fusion protein occurs as a result of RNA splicing.

CLIN. CORR. 18.3 THE THALASSEMIAS

In the diseases known as the thalassemias the amount of a particular globin chain is severely reduced or absent. There is a deficiency of α-globin chains in α-thalassemia, with a resultant precipitation of β-chains. Homozygotes with α^0-thalassemia have a syndrome called *hydrops fetalis* in which death occurs prior to or shortly after birth. The genes for the α chain are on chromosome 16, while genes

for many of the other globin types are on chromosome 11.

In the β-thalassemias the amount of β-globin varies. No β-globin mRNA can be detected in one type of β^0-thalassemia. In another, mRNA is present in the nucleus but not the cytoplasm. In a third type the β-globin mRNA is present but not translated, while α-globin mRNA is translated normally. β^+-thalassemia is characterized by low concentrations of cellular β-globin chains. These are translated from reduced amounts of mRNA. Both β- and δ-globins are absent in $\delta\beta^0$-thalassemia. This condition is due to a deletion from the 5′ region of the δ-globin gene to the 3′ side of the β-globin gene (Figure 18.22).

chickens. The ovalbumin mRNA has 1,890 nucleotides and is processed from a primary transcript of 7,900 nucleotides having a cap and poly(A) tail. The primary transcript has a 5′ leader sequence of 76 nucleotides, 7 mRNA exons (185, 51, 129, 118, 143, 156, and 1,043 nucleotides), and 7 introns (1,560, 238, 601, 411, 1,029, 323, and 1,614 nucleotides) (see Figure 18.20).

Another chicken protein, ovomucoid, is translated from a small mRNA of 800 nucleotides. However, this mRNA is processed from a primary transcript having 6,000 nucleotides and at least six intervening sequences.

Histone mRNAs are also processed prior to translation. Nuclear pre-mRNAs of about 14,000 nucleotides have been found, that do not have intervening sequences and therefore do not require splicing. The histone mRNAs have 5′ caps, $m^7G^{5'}ppp^{5'})X^mpY$ . . ., but only some have poly(A) tails. Each histone mRNA contains noncoding leader and trailer sequences. Histones are found in all eucaryotic cells and have resisted evolutionary pressures. The genes for the five major histones are clustered within the genome in a repetitive and tandemly linked arrangement. There are 30 to 40 copies of each histone gene in human cells and each gene is linked by a variable spacer region having a high percentage of A · T base pairs. Histone mRNA translation is linked, and transcription of histone genes also may be linked, to DNA replication in the S phase of the cell cycle.

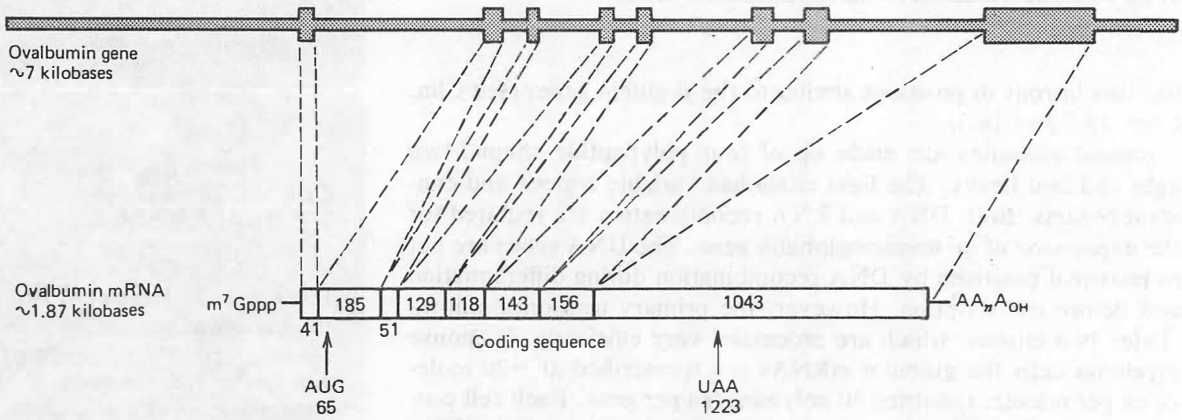


Figure 18.20
Diagram of the ovalbumin gene or primary transcript and the several RNA splicing sites required to synthesize the ovalbumin mRNA.
The two arrows indicate the translated portion of the mRNA.
Redrawn with permission from P. Chambon et al., *Miami Winter Symp.*, 16:57, 1979.

18.6 NUCLEASES

The nucleases are a group of enzymes that cleave the phosphodiester bonds in polynucleotides. There are three types of nucleases for RNA: the ribonucleases (RNase), which are RNA-specific and do not hydrolyze DNA; the phosphorylases, such as polynucleotide

Table 18.6 Characteristics of Ribonucleases

Nuclease	*Specificity*	*Products*
Endonucleases yielding 3′-phosphates		from 5′-pApGpCpGpUpU$_{OH}$-3′
Pancreatic RNase	After pyrimidines	pApGpCp + GpUp + U$_{OH}$
T_1	After guanines	pApGp + CpGp + UpU$_{OH}$
U_1	After guanines	pApGp + CpGp + UpU$_{OH}$
T_2	After adenines	pAp + GpCpGpUpU$_{OH}$
Rat liver RNase-1	All phosphodiester bonds	pAp + 2Gp + Cp + Up + U$_{OH}$
Rat liver RNase-2	Between adjacent pyrimidines	pApGpCpGpUp + U$_{OH}$
E. coli RNase III	Double-helical structures	Cleaves rRNA precursors
Endonucleases yielding 5′-phosphates		
Rat liver alkaline RNase I	Nonspecific	pA + 2pG + pC + pU + pU$_{OH}$
E. coli RNase P also in mammalian cells	Precursor tRNA	Mature tRNA + fragment
E. coli RNase H also in mammalian cells	RNA of DNA–RNA hybrid	Nucleoside 5′-monophosphates + DNA
Exonucleases		
E. coli RNase II also in mammalian cell nuclei	Single strands, 3′ ⟶ 5′	Nucleoside 5′-monophosphates, may trim RNA precursors after endonucleases
E. coli RNase V	Single strands, 5′ ⟶ 3′ precursor mRNAs	Nucleoside 5′-monophosphates
E. coli oligoribonuclease	Short oligoribonucleotides	Nucleoside 5′-monophosphates
RNase H from RNA tumor virus	RNA of DNA · RNA hybrid 5′ ⟶ 3′ 3′ ⟶ 5′	Nucleoside monophosphates
Polynucleotide phosphorylase	Single-strand RNA 3′ ⟶ 5′	Nucleoside 5′-diphosphates
Nonspecific nucleases		
Micrococcal endonuclease	Single-strand RNA or DNA	Nucleoside 3′-monophosphates
Nuclease S1	Single-strand RNA or DNA	5′-Phosphate oligonucleotides
Venom phosphodiesterase	Exonuclease 3′ ⟶ 5′ Blocked by 3′-phosphate end	Nucleoside 5′-monophosphates
Spleen phosphodiesterase	Exonuclease 5′ ⟶ 3′	Nucleoside 3′-monophosphates

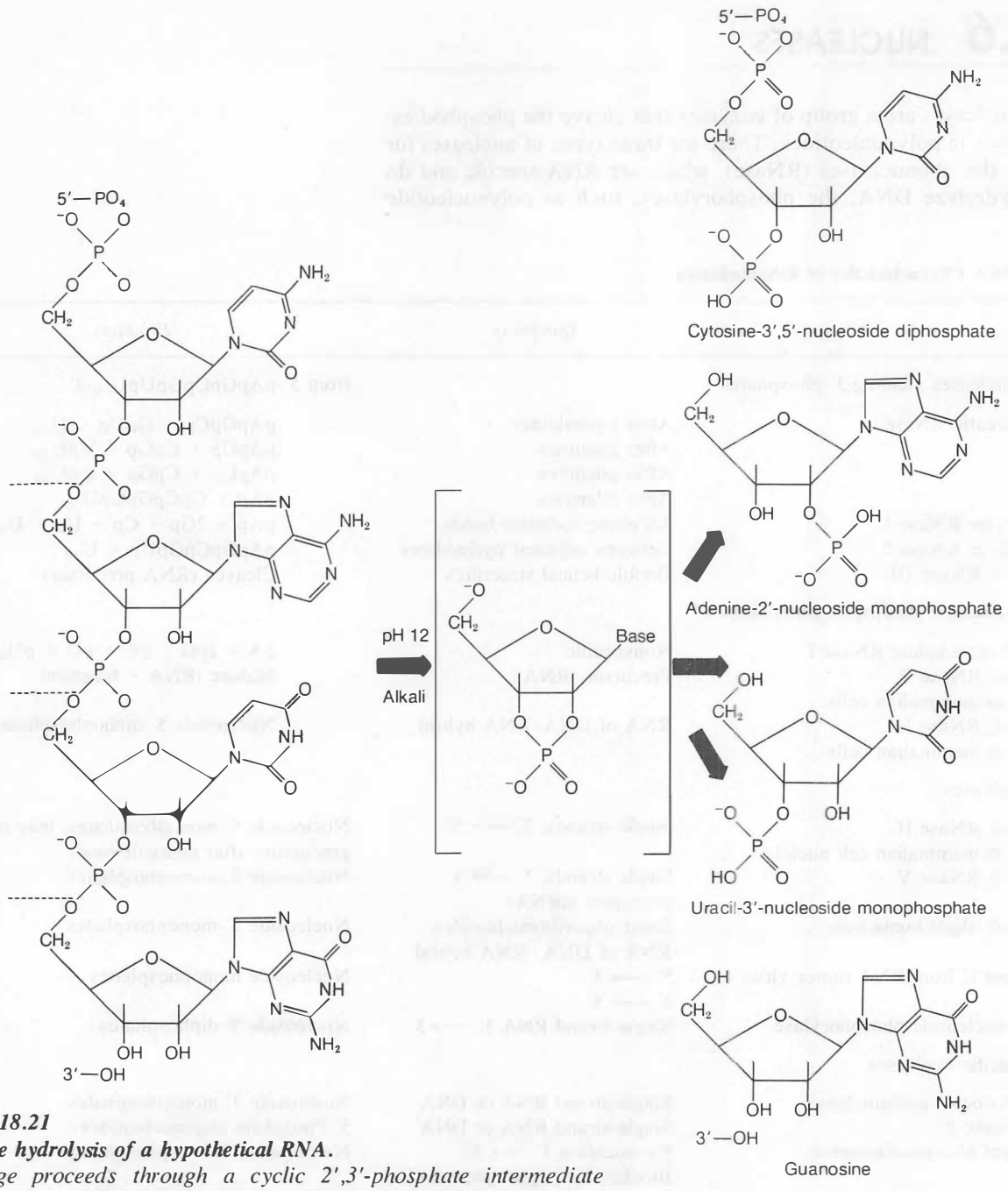


Figure 18.21
Alkaline hydrolysis of a hypothetical RNA.
Cleavage proceeds through a cyclic 2′,3′-phosphate intermediate and results in a 3′,5′-diphosphate nucleoside from the 5′ terminus, 2′- or 3′-nucleoside monophosphates from the internal residues and a nonphosphorylated nucleoside from the 3′ terminus. DNA is not subject to alkaline hydrolysis.

phosphorylase and pyrophosphorylase; and the phosphomonoesterases. These enzymes show specificities for secondary structure, direction of attack, and bond cleavage. Unlike the restriction endonucleases for DNA, the RNases are not sequence-specific; however, their different modes of action have made the RNases invaluable tools in RNA nucleotide sequence analysis. The most useful for sequence studies have been the pancreatic, T_1, and T_2 RNases.

Ribonucleases may be either endonucleases or exonucleases. Table 18.6 summarizes the representative nucleases.

Endonucleases form ribonucleoside monophosphates and oligonucleotide fragments having phosphates on the 3′ or 5′ positions of the ribose. Each endonuclease is specific in this manner. Some, such as pancreatic RNase, cleave only after the pyrimidines uracil and cytosine, yielding only pyrimidine nucleotides and fragments with 3′-terminal pyrimidines and 5′-purines. Other RNases, T_1 and T_2, cleave only after specific purines yielding free purine nucleotides and fragments with 3′-terminal purines and 5′-pyrimidines. The RNA endonucleases generally cleave only single-stranded or nonpaired regions of RNA. There are a few RNases used during maturation of precursor tRNAs and mRNAs, which cut near or in base-paired regions. Proteins that bind to RNA, such as ribosomal proteins, also protect RNA from nucleolytic attack. RNA with specific binding proteins attached can be treated with RNases, leaving only the small RNA fragment bound to the protein. This oligonucleotide can be sequenced to determine the actual protein recognition and binding site on the RNA.

The exonucleases attack RNA from either end and work sequentially along the RNA, yielding 3′- or 5′-nucleoside monophosphates. Each exonuclease, except RNase H, functions in a unique direction and usually removes only non-base-paired nucleotides. Polynu-

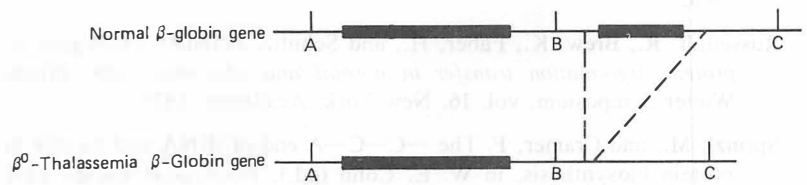


Figure 18.22
Diagram showing the deleted region of the β-globin gene in β° thalassemia.
A, B, and C indicate markers in the DNA. A 600-base pair segment is deleted from the normal gene.

cleotide phosphorylase has the dual capacity of polymerizing nucleotides to form oligonucleotides or depolymerizing RNA in the presence of inorganic phosphate to yield nucleoside 5′-diphosphates.

Treatment of RNA with alkali (e.g., KOH or NaOH, pH 12) causes breaks in the sugar-phosphate backbone, resulting in nucleoside 2′- or 3′-monophosphates (Figure 18.21). In addition, there will be one nucleoside (2′)3′,5′-diphosphate representing the 5′ terminus of each RNA and one nucleoside from the 3′ terminus of each RNA molecule. This hydrolysis occurs only with RNA, not DNA, because of the 2′-OH group, which permits the 2′-3′ cyclic intermediate to form. Methylation of ribose at the 2′-carbon blocks the lytic action of alkali.

BIBLIOGRAPHY

Abelson, J. RNA processing and the intervening sequence problem. *Ann. Rev. Biochem.* 48:1035, 1979.

Cohn, W. E. (ed.). *Progress in nucleic acid research and molecular biology*, vol. 23. New York: Academic, 1979.

Cohn, W. E., and Volkin, E. (eds.). *Progress in nucleic acid research and molecular biology, mRNA: the relation of structure to function*, vol. 19. New York: Academic, 1976.

Davidson, J. N. *The biochemistry of the nucleic acids*, 8th ed., revised by R. L. P. Adams. New York: Academic, 1976.

Losick, R., and Chamberlain, M. (eds.). *RNA polymerase*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1976.

Rosenberg, M., and Court, D. Regulatory sequences involved in the promotion and termination of RNA transcription. *Ann. Rev. Genet.* 13:319, 1979.

Russell, T. R., Brew, K., Faber, H., and Schultz, J. (eds.). *From gene to protein: information transfer in normal and abnormal cells*, Miami Winter Symposium, vol. 16. New York: Academic, 1979.

Sprinzl, M., and Cramer, F. The —C—C—A end of tRNA and its role in protein biosynthesis, in W. E. Cohn (ed.), *Progress in nucleic acid research and molecular biology*, vol. 22. New York: Academic, 1979, p. 1.

19

Protein Synthesis

KARL H. MUENCH

TRANSLATION AND POSTTRANSLATIONAL MODIFICATION

19.1 THE GENETIC CODE

Information necessary for synthesis of the unique primary structure of each protein is analogous to the written human languages that consist of a sequence of letters forming words and sentences. In the genetic language, however, there are only four letters. Four nucleotidyl residues, differing only in the four purine and pyrimidine base residues, A, G, U, C (adenine, guanine, uracil, cytosine), recur in various permutations of sequence embodied in mRNA. The concept of a code arises from the need for the nucleotidyl residue sequences, hereafter designated base sequences, to be translated into amino acid residue sequences, constituting proteins. Specific contiguous base sequences define the code message for each of the 20 coded amino acids found in proteins. Thus in translation a four-letter language becomes a 20-letter language. Implicit in the analogy to language is the necessity for directional specificity. Base sequences are by convention written left to right from the 5′ terminal to the 3′ terminal, and amino acid sequences are by convention written left to right from the amino terminal to the carboxy terminal. Later we shall see that these directions in mRNA and proteins correspond.

Codons Are Three-Letter Words

The four bases taken as pairs give only 16 permutations, as shown in Table 19.1. This hypothetical code of two-letter words is illustrative but would be insufficient to specify 20 different amino acids, not to mention start and stop signals. Actually, the genetic code consists of three-letter words. The four bases taken by threes give 64 permutations (words), more than sufficient to designate 20 amino acids. These words of three bases are called codons. The 64 codons are customarily arranged in tabular form, as shown in Table 19.2, and comprise the genetic code. Perusal of the genetic code reveals that different amino acids have different numbers of codons. For example, tryptophan has only one codon, whereas arginine, leucine, and serine each have six codons. The existence of multiple codons for one amino acid has been called "degeneracy."

A useful way to review the tabular genetic code is as a composite of 16 boxes, each containing four codons. In each box of four codons the first two bases are the same. In 8 of the 16 boxes is seen a four-codon family for a single amino acid. Members of these codon families differ only in the third base. In these eight codon families two-base codons would seem adequate, and there is evidence that in some cases only the first two bases are sufficient to specify the amino acid. We shall return to this point below.

Table 19.1 Sixteen Permutations of a Four-Letter Code Taken by Pairs

First Base	*Second Base*			
	U	*C*	*A*	*G*
U	UU	UC	UA	UG
C	CU	CC	CA	CG
A	AU	AC	AA	AG
G	GU	GC	GA	GC

Table 19.2 The Genetic Code[a]

5′ Base	U		C		A		G		3′ Base
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG		UGG	Trp	G
C	CUU		CCU		CAU	His	CGU		U
	CUC	Leu	CCC	Pro	CAC		CGC	Arg	C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU		ACU		AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU		GCU		GAU	Asp	GGU		U
	GUC	Val	GCC	Ala	GAC		GGC	Gly	C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

[a] The genetic code comprises 64 codons, which are permutations of four bases taken by threes. Note the importance of sequence: three bases give six permutations (codons), ACG, AGC, GAC, GCA, CAG, and CGA, for threonine, serine, aspartate, alanine, glutamine, and arginine, respectively.

Punctuation Codons

Four of the 64 codons specify start and stop signals for protein synthesis. The start signal, AUG, also specifies the amino acid methionine. When AUG appears at the beginning of a message, methionine becomes the initial amino acid at the amino terminal of the protein to be synthesized. Because AUG is the only methionine codon, it also specifies methionine residues in the interior of the protein sequence. In contrast to the start signal, the stop signals, UAA, UAG, and UGA, specify no amino acid. For that reason they are sometimes called nonsense codons.

Codon–Anticodon Interaction

With the partial exception of unique codon usage in mitochondria, the genetic code is universal. We shall see that the code as deciphered from studies using bacterial components is precisely the same as that used by human beings for synthesis of all proteins. The translation of a codon is accomplished by a mechanism involving

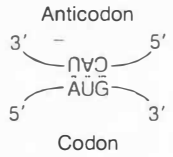


Figure 19.1
The codon–anticodon interaction for methionine.

complementary anticodons. The codon–anticodon interaction for methionine is shown in Figure 19.1. The anticodons are three-base sequences in tRNA, as discussed in Chapter 18. When tRNAs are depicted in their secondary or cloverleaf structures, as shown in Figure 18.4, the anticodons are always found in the anticodon loop, remote from the amino acid acceptor stem. Anticodon sequences in tRNAs of known specificity have been used to confirm the genetic code. Conversely, tRNAs of unknown amino acid specificity have been identified by their anticodons. When tRNA is depicted in its L-shaped tertiary structure as in Figure 18.4, the anticodon is at one extremity of the molecule and the amino acid residue attachment point on the 3′ terminus of the acceptor stem is at the other extremity of the molecule. Thus the structure of tRNA is ideally suited to its role as an adaptor or bridge between the base sequence in mRNA and the amino acid sequence in nascent protein.

For reasons that are not understood, mitochondria use certain codons differently. The codons involved in this unique isolation of the mitochondrion from the genetic code are shown in the table.

Codon	*General Code*	*Mitochondrial Code*
CUA	Leu	Thr
AUA	Ile	Met
UGA	Stop	Trp

Exceptions to Classical Base Pairing

Perfect adherence to Watson–Crick base pairing would demand absolute specificity of a given tRNA for a given codon. However, the degeneracy of the genetic code is not matched by a precise correspondence in multiplicity of tRNAs, and one tRNA may actually translate several different codons, all with the same bases in the first two positions. Exceptions to Watson–Crick base pairing at the third position of the codon are given in the "wobble" rules, according to which G in the third position of the codon may pair with U in the first position of the anticodon, U in the third position of the codon may pair with G in the first position of the anticodon, and U, C, or A in the third position of the codon may pair with I (hypoxanthine) in the first position of the anticodon:

Anticodon	3′	– – U	– – G	– – I	– – I	– – I
Codon	5′	– – G	– – U	– – U	– – C	– – A

Hypoxanthine is one of the numerous unusual bases found in tRNA. Some of these unusual bases are found in or near anticodons. The wobble rules are not sufficient to explain other non-Watson–Crick codon–anticodon interactions that do occur. For example, in the case of valine all four codons (GUU, GUC, GUA, and GUG) can be read by each of the anticodons U*AC, GAC, and IAC (where U* is the unusual base, 5-oxyacetic acid uridine). Therefore, either base pairs prohibited by wobble rules (I : G, G : G, and A : G) actually do form, or a two-letter code suffices for translation in this valine family. According to the wobble rules 31 different tRNAs would suffice to read the 64 codons. However, human (HeLa cell) mitochondria contain only 23 tRNAs, which must then be translating certain codons by the first two letters.

Deciphering the Code

In the presence of higher than physiological concentrations of magnesium ion in vitro, specific initiation components required for protein synthesis in vivo are not necessary, and the synthetic polymers, poly(A), poly(U), and poly(C), are readily translatable to produce the synthetic polypeptides, polylysine, polyphenylalanine, and polyproline, respectively. From such studies one can deduce the respective codons to be AAA, UUU, and CCC. Interesting variations and extrapolations of such experiments led historically to the complete deciphering of the genetic code. For example, a perfectly alternating copolymer, AUAUAUAUAUAUAUAU. . . or $(AU)_n$ contains two and only two codons, AUA and UAU. Furthermore, these two codons occur not at random but in a perfectly alternating sequence, a fact leading to the expectation that translation of this synthetic messenger would produce a polypeptide, Ile-Tyr-Ile-Tyr-Ile-Tyr-, and this was indeed realized. Similarly a perfect synthetic message containing the sequence $(CUG)_n$ contains the possible codons CUG for leucine, UGC for cysteine, and GCU for alanine. However, in this case any one codon, once selected, sets the reading frame and is followed by repetition of the same codon. The selection of a particular initial codon in this case is random. We shall later see that selection of the initial codon for protein synthesis in vivo is not random but specific. A special initiation codon, the start signal AUG, sets the unique reading frame for most natural mRNAs. The synthetic proteins realized from the synthetic message $(CUG)_n$ are three different homogeneous polypeptides: polyleucine, polycysteine, and polyalanine. The student should here be convinced that a perfectly regular sequence of, for example, (CUCG) would produce a single synthetic polypeptide of the sequence (Leu-Ala-Arg-Ser). These data are summarized in Table 19.3. The results show that in

CLIN. CORR. 19.1 MISSENSE MUTATION: HEMOGLOBIN

Clinically the most important missense mutation known is the change from A to U in either the GAA or GAG codon for glutamate to give a GUA or GUG codon for valine in the sixth position of the β chain for hemoglobin. An estimated 1 of 10 American blacks are carriers of this mutation, which in its homozygous state is the basis for sickle cell disease, the most common of all hemoglobinopathies. The second most common hemoglobinopathy is hemoglobin C disease, in which a change from G to A in either the GAA or GAG codon for glutamate results in an AAA or AAG codon for lysine in the sixth position of the β chain.

CLIN. CORR. 19.2 DISORDERS OF TERMINATOR CODONS

In the disorder of hemoglobin McKees Rocks the UAU or UAC codon normally designating tyrosine in position 145 of the β chain has mutated to the terminator codon UAA or UAG. The result is a shortening of the β chain from its normal 146 residues to 144 residues, a change that gives the hemoglobin molecule an unusually high oxygen affinity. Another human illness resulting from a terminator mutation is a variety of β-thalassemia. The thalassemias comprise a group of disorders characterized at the molecular level

Table 19.3 Polypeptide Products of Synthetic mRNAs[a]

mRNA	*Codon Sequence*	*Products*
—$(AU)_n$—	—AUA UAU AUA UAU—	—$(\text{Ile-Tyr})_{n/3}$—
—$(CUG)_n$—	—CUG CUG CUG CUG—	—Leu_n—
	—UGC UGC UGC UGC—	—Cys_n—
	—GCU GCU GCU GCU—	—Ala_n—
—$(CUCG)_n$—	CUC GCU CGC UCG	—$(\text{Leu-Ala-Arg-Ser})_{n/3}$—

[a] The horizontal brackets accent the reading frame.

protein synthesis the codons are taken in their precise sequence without the omission of a single base and without the double use of any base. In the latter case the reading frames would be overlapping rather than sequential. Such an overlapping reading frame sequence would place restrictions on which amino acids could follow others and would allow a single base change to alter three resulting amino acid residues instead of only one. The absence of these phenomena proves the nonoverlapping nature of the genetic code.

Mutations

The genetic code provides a basis for understanding mutation. Originally a genetic term, the word *mutation* describes a change in a gene, and some mutations have been characterized in biochemical terms. For example, missense mutations involve a codon change, usually by a single base, such that the mutant codon designates a different amino acid than does the original codon. Of almost 200 mutations expressed in the α chain of human hemoglobin the vast majority are of the missense variety. Every one of these missense mutations confirms the genetic code, because the amino acid replacement in the mutant hemoglobin is reducible to and explained by a single base change in a codon (Clin. Corr. 19.1).

Just as a codon may change to designate another amino acid, so an amino acid codon may change to a stop codon, thereby producing a terminator mutation. Two human illnesses so far are known to result from terminator mutations. One is a polycythemia, resulting from the extremely high oxygen affinity of hemoglobin McKees Rocks, and the other is β-thalassemia (Clin. Corr. 19.2).

When a stop codon mutates to become a sense codon, then base sequences not normally translated at the 3′ end of the mRNA are

Table 19.4 Reverse Terminator Mutations Producing Abnormal α-Globins

Hemoglobin	*α-Codon 142*	*Amino Acid 142*	*α-Globin Length (residues)*
A	UAA	—	141
Constant Spring	CAA	Glutamine	172
Icaria	AAA	Lysine	172
Seal Rock	GAA	Glutamate	172
Koya Dora	UCA	Serine	172

read as sequential codons until the appearance of another stop codon. The result is the appearance of a larger protein in place of the normal protein. This phenomenon is the basis for the several human disorders summarized in Table 19.4 and discussed in Clin. Corr. 19.3. Although all the data in Table 19.4 are equally consistent with UAG as the normal stop codon for α-globin, evidence from another kind of mutation, a frameshift mutation, indicates that UAA is indeed the normal stop codon. Frameshift mutations result from a single base deletion or addition and may cause the appearance of a sequence of different amino acids or the appearance or disappearance of a stop mutation. Hemoglobin Wayne contains an abnormal α-globin chain as a consequence of a frameshift mutation, illustrated in Table 19.5. A deletion of the indicated U in codon 138 causes the resulting sequential codon changes beginning at the next position, 139, such that the amino acid sequence changes. In addition, the normal stop codon UAA is now no longer read in that frame, and protein synthesis continues beyond that point until another stop codon in the new reading frame is encountered at position 147 as shown. If UAG were the normal stop codon for α-globin, as would be allowed from the data in Table 19.4, then codon 142 in Wayne α-globin would be AGG for arginine instead of AAG for lysine as actually found.

by an inbalance in the stoichiometry of globin synthesis. In β^0-thalassemia no β-globin is synthesized. As a result α-globin, unable to associate with β-globin to form hemoglobin, accumulates and precipitates in erythroid cells. The precipitation damages cell membranes, causing hemolytic anemia and interfering with erythropoiesis. One variety of β^0-thalassemia results from a terminator mutation at codon 17 of the β-globin. The normal codon AAG designates a lysyl residue at β-17 but becomes the stop codon UAG in this variety of β^0-thalassemia. In contrast to the situation with hemoglobin McKees Rocks, in which the terminator mutation occurs late in the β-globin message, the terminator mutation occurs so early in the mRNA of β^0-thalassemia that no useful partial β-globin sequence can be synthesized, and β-globin is absent.

CLIN. CORR. 19.3 THALASSEMIA

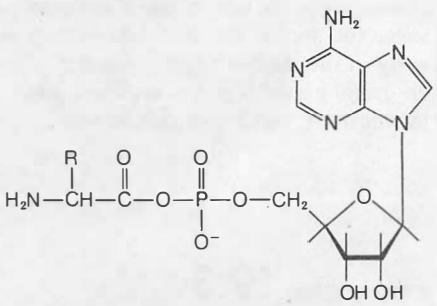
The disorders summarized in Table 19.4 are forms of α-thalassemia, because the abnormally long α-globin molecules, which replace normal α-globin, are present only in small amounts. The small amounts

Table 19.5 A Human Frameshift Mutation Produces Hemoglobin Wayne

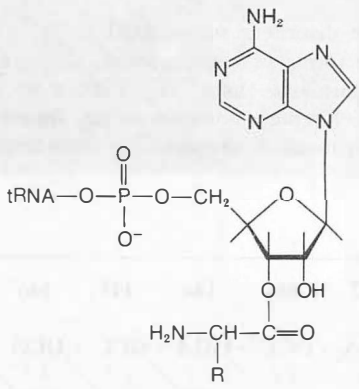
Position	137	138	139	140	141	142	143	144	145	146	147
Normal α-globin amino acid sequence	- Thr	- Ser	- Lys	- Tyr	- Arg						
Normal α-globin codon sequence	- ACP	- UC(U)	- AAA	- UAC	- CGU	- [UAA]	- GCU	- GGA	- GCC	- UCG	- GUA
Wayne α-globin codon sequence	- ACP	- UCA	- AAU	- ACC	- GUU	- AAG	- CUG	- GAG	- CCU	- CGG	- [UAG]
Wayne α-globin amino acid sequence	- Thr	- Ser	- Asn	- Thr	- Val	- Lys	- Leu	- Glu	- Pro	- Arg	

The base deletion causing the frameshift is encircled. The stop codons are boxed.
P = A, G, U, or C.

of α-globin result either from decreased rate of synthesis or from increased rate of breakdown. As can be seen, the normal stop codon, UAA, for α-globin mutates to any of four sense codons with resultant placement of four different amino acids at position 142. Normal α-globin is only 141 residues in length, but the four abnormal α-globins are 172 residues in length, presumably because of a terminator codon in position 173.



Aminoacyladenylate



Aminoacyl-tRNA

Figure 19.2
Intermediates in amino acid incorporation into protein.

The importance of reading frame is underscored by the phenomenon of overlapping genes. Although the genetic code is not overlapping, certain viruses use the same length of DNA to encode different genes in different reading frames. In this way more information can be stored in the DNA, which for reasons of viral packaging must be limited in total amount. Thus the genome of ϕX174 is too small to code for its nine proteins without use of different and overlapping reading frames. At one point of some viral DNA all three reading frames may be used. Mammalian viruses with overlapping genes include simian virus 40, which causes tumors in apes, and the closely homologous BK virus, isolated from human sources.

19.2 AMINOACYL-tRNA SYNTHETASES

Protein synthesis involves convergence of diverse components. A useful way to conceptualize the process as a sequence of reactions is to follow the pathway of amino acids. From the "point of view" of amino acids, the aminoacyl-tRNA synthetases catalyze the first two steps in protein synthesis. These are the activation of each amino acid as aminoacyl-AMP, a mixed acid anhydride (Figure 19.2) and the subsequent attachment of each amino acid to its cognate tRNA in ester linkage (Figure 19.2) for later transfer into polypeptide. The two steps may be written with *L*-valine, for example,

$$\text{L-Valine} + \text{ATP} + [\text{ValRS}] \rightleftharpoons \text{Val} \sim \text{AMP} \cdot [\text{ValRS}] + \text{PP}_i$$

$$\text{Val} \sim \text{AMP} \cdot [\text{ValRS}] + \text{tRNA}^{\text{Val}} \rightleftharpoons$$

$$\text{Val} \sim \text{tRNA}^{\text{Val}} + \text{AMP} + [\text{ValRS}]$$

Valyl-tRNA synthetase (ValRS) is entered as a reactant because of the formation of Val~AMP, not as a free intermediate but as an enzyme-bound intermediate. The sum of the two reactions is then

$$\text{L-Valine} + \text{ATP} + \text{tRNA}^{\text{Val}} \rightleftharpoons \text{Val} \sim \text{tRNA}^{\text{Val}} + \text{AMP} + \text{PP}_i$$

Viewed in this way ATP is split to AMP and PP_i with concomitant formation of one aminoacyl-tRNA. The actual equilibria of these reactions in vivo is shifted far to the product side because of cleavage of pyrophosphate by ubiquitous pyrophosphatases.

Each of the 20 coded amino acids has a specific aminoacyl-tRNA synthetase, and each aminoacyl-tRNA synthetase recognizes one or several cognate tRNAs, capable of linking only the specific amino acid. The specificity of aminoacyl-tRNA biosynthesis is as important for accurate protein biosynthesis as is the specificity of the codon–anticodon interaction. This is true because once attached to tRNA an amino acid residue has no effect on its own specific placement into protein, that placement now becoming entirely the responsibility of the codon–anticodon interaction. Thus in vitro cysteinyl-$tRNA^{Cys}$ can be reduced with Raney nickel to alanyl-$tRNA^{Cys}$. The chemically modified aminoacyl-tRNA has been used as a precursor in hemoglobin biosynthesis by reticulocyte lysates, and the alanine residues occupy known positions of cysteine, but not of alanine, in the finished hemoglobin.

Mechanisms to Insure Fidelity

Accurate protein synthesis depends absolutely on mechanisms to insure fidelity of amino acid attachment to tRNA. These mechanisms are remarkably discriminating. For example, consider the minimal structural difference, a single methylene group, between valine and isoleucine. During the synthesis of rabbit hemoglobin, even the most difficult error to prevent, the substitution of valine for isoleucine, occurs only about one in 3,000 opportunities. Such fine discrimination cannot be explained only by relative K_m values in aminoacyl-tRNA synthesis. Then what is the explanation? The double-sieve mechanism is a useful model hypothesis. This hypothesis divides the problem into two parts: discrimination against larger analogs and discrimination against smaller analogs. For example, take ValRS and isoleucine for a model of discrimination against a larger analog. As shown in Figure 19.3 ValRS discriminates against isoleucine in the activation step. The relative rates of activation of valine and of isoleucine depend on their concentrations and on their K_{cat}/K_m ratios. Thus

$$v = \frac{K_{cat}}{K_m}[E][S]$$

at any substrate concentration, where v is the reaction rate, [E] is concentration of free enzyme, and K_{cat} is the turnover number. The reaction rates for valine and for isoleucine can then be compared according to the equation:

$$\frac{v_{val}}{v_{Ile}} = \frac{[Val](K_{cat}^{Val}/K_m{}^{Val})}{[Ile](K_{cat}^{Ile}/K_m{}^{Ile})}$$

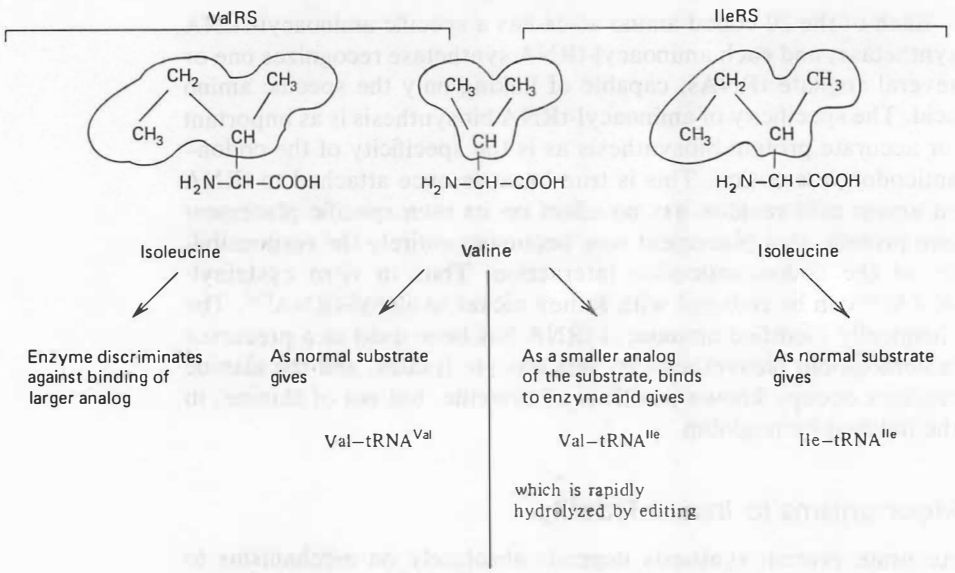


Figure 19.3
Discrimination against larger and smaller analogs by the double sieve mechanism.

The ratio actually measured in model bacterial systems is about 10^5, sufficient to explain the observed mistake levels in vivo.

Now take IleRS and valine for a model of discrimination against an isosteric or a smaller analog. Although IleRS discriminates against valine, the relative rates of valine and isoleucine activation are not different enough to explain the low, observed in vivo mistake levels. In this case the second reaction occurs, and Val-tRNA^{Ile} is actually formed. However, a hydrolytic site on the IleRS surface recognizes this mismatch and rapidly hydrolyzes the ester bond to give valine and tRNA^{Ile}. This editing function is analogous to the 3′ to 5′ editing function of DNA polymerase I.

19.3 PROTEIN SYNTHESIS

The needs of various cells to synthesize protein vary markedly. At one extreme are mature red blood cells, which lack the enzymes and organelles required for protein synthesis. They undergo no cell divi-

sion and have a finite life span of about 120 days because of their inability to replenish necessary proteins. Other cells must maintain levels of certain enzymes and replace structural proteins to remain viable. For growing and dividing cells greater levels of protein synthesis are required. Finally, some cells synthesize relatively large amounts of protein for export. Examples are pancreatic acinar cells, which synthesize digestive enzymes such as trypsin and chymotrypsin, endocrine cells such as the insulin-producing β cells of the pancreas, and liver cells, which synthesize many proteins, including serum albumin. These cells with the highest protein synthetic activities contain large numbers of ribosomes, which lend the cytoplasm its basophilic staining qualities, seen by light microscopy.

In essence protein synthesis involves the convergence of sequence information in the form of mRNA, activated amino acids in the form of aminoacyl-tRNAs, energy in the form of GTP, and various protein factors. The process occurs on the surface of ribosomes, which provide at least one required enzyme. That enzyme is peptidyltransferase, one of many proteins making up the larger ribosomal subunit and imbedded in the surface of that subunit. Peptidyltransferase catalyzes peptide bond formation, the actual covalent linkage of one amino acid residue to another. The rate of this bond formation can be as high as 1,200 residues per minute per ribosome at 37°C. The rate of peptide bond formation per mRNA molecule is much higher because many ribosomes simultaneously translate each mRNA. The resulting polyribosomes are visible by electron microscopy, as shown in Figure 19.4, and are measurable by zone sedimentation in sucrose gradients as shown in Figure 19.5.

Structure of mRNA

In contrast to prokaryotes mRNA in eukaryotes is synthesized in the nucleus and must cross the nuclear membrane to be translated in the cytoplasm. This mRNA transport is closely related to the mRNA processing described in Chapter 18. Thus eukaryotic mRNA is complex and contains more than the simple base sequence required to specify a protein. To review briefly, at the 5′ end is a characteristic cap structure. After an untranslated region near the 5′ end is a ribosome binding site consisting of about 10 base residues believed to pair with a complementary sequence in ribosomal RNA. Shortly after the ribosomal binding site is the AUG codon and the base sequence for the protein, followed by a stop codon. Eukaryotic mRNAs are usually monocistronic, that is, carrying the base sequence for only a single protein. After the stop codon is a noncoding sequence before the 3′ terminus of poly A. The noncoding sequence is at least 96 base residues long in mRNA for human α-globin, as

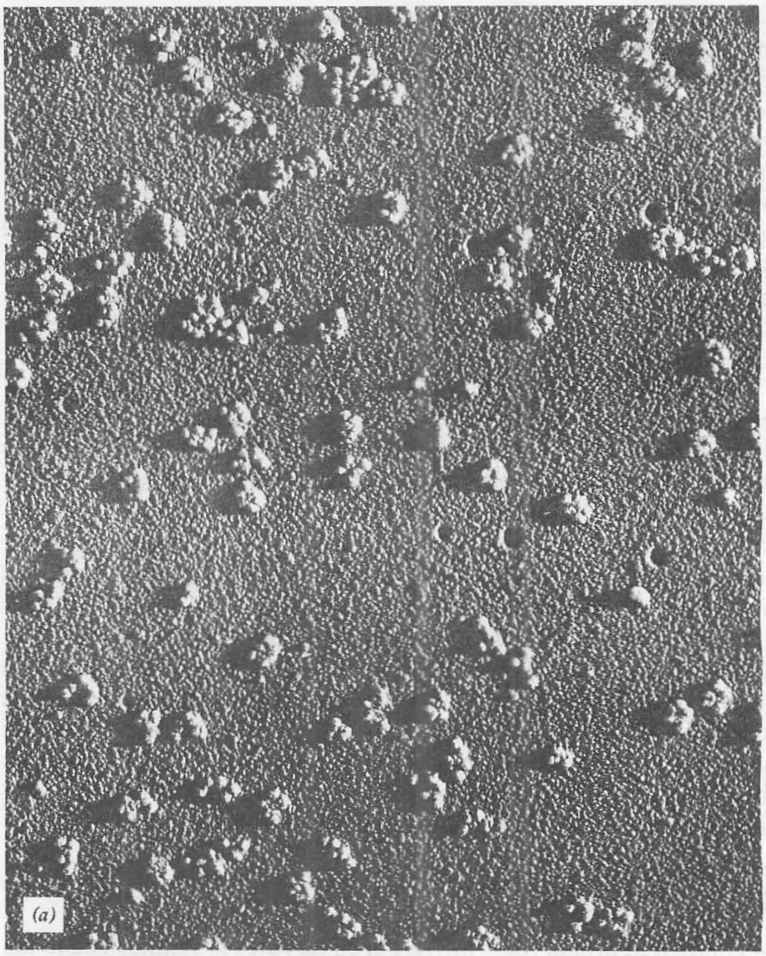


Figure 19.4
Reticulocyte polyribosomes shadowed with platinum as shown at left appear as clusters of three to six ribosomes, a number consistent with the size of mRNA for a globin chain.
Further magnification after uranyl acetate staining as shown on the right reveals one extraordinarily clear five-ribosome polysome with part of the mRNA visible.
Courtesy of Dr. Alex Rich, MIT.

shown by the chain termination mutants already discussed: those giving hemoglobins Constant Spring, Icaria, Seal Rock, and Koya Dora (Table 19.4). This noncoding sequence must have an important normal function, because its sequence is highly conserved, being 80% homologous from rabbit to man.

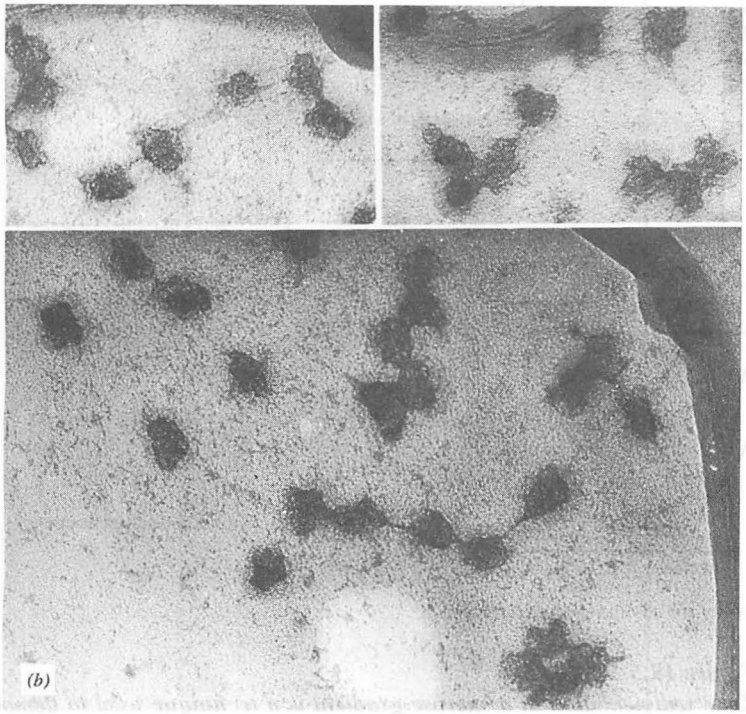


Figure 19.4 (Continued)

Direction of Translation

Just as these words are read in sequence from left to right, so mRNA is translated in a defined direction and not in a random fashion. The defined direction of translation is such that the amino terminal of the nascent protein is synthesized first, and the carboxy terminal of the nascent protein is synthesized last. This can be demonstrated in reticulocytes synthesizing hemoglobin at artificially lowered temperature so as to decrease the rate of protein synthesis and make the process more amenable to study. In such a system radioactive amino acids are added for short, measured periods of time, and the free globin chains synthesized are then isolated for analysis. As shown in Figure 19.6 the radioactive precursor should be found in the portion of the globin chain synthesized last. After longer and longer exposure to the radioactive precursor more and more of the globin chain becomes labeled until eventually even the first part of the molecule to be synthesized is labeled with radioactivity. The peptides obtained by digesting the globin product are arranged in order of specific radioactivity. From the known sequence of the peptides in

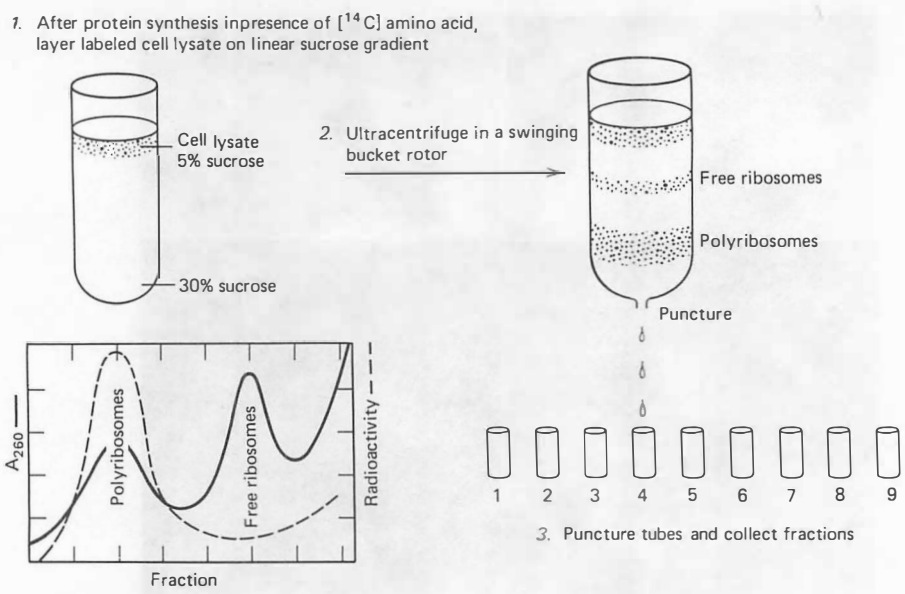


Figure 19.5

Zone sedimentation in a sucrose gradient is a technique used to demonstrate that polyribosomes, not free ribosomes, are the site of protein synthesis.

The four basic steps of the technique are indicated. Polyribosomes labeled with [^{14}C]amino acid in a cell lysate actively synthesizing protein sediment in the radioactive peak. The absorbance at 260 nm measures RNA, mostly rRNA with a prominent peak at the position of free ribosomes.

globin it can be seen that the most highly radioactive peptides, for example, peptides F and G in Figure 19.6, are always at the carboxy end of the globin. Conversely, the least radioactive peptide, peptide A in Figure 19.6, is always at the amino end of the globin. Therefore the temporal sequence of protein synthesis is from the amino to the carboxy end.

Information on the direction of protein synthesis and knowledge of the genetic code together prove that the direction of translation of mRNA must be 5′ to 3′. Thus mRNA is translated in the same direction in which it is transcribed, and the 5′ end of mRNA corresponds to the amino end of the protein for which the mRNA codes. The existence of stable, isolatable polyribosomes and the directional, sequential nature of protein synthesis imply that ribosomes do not

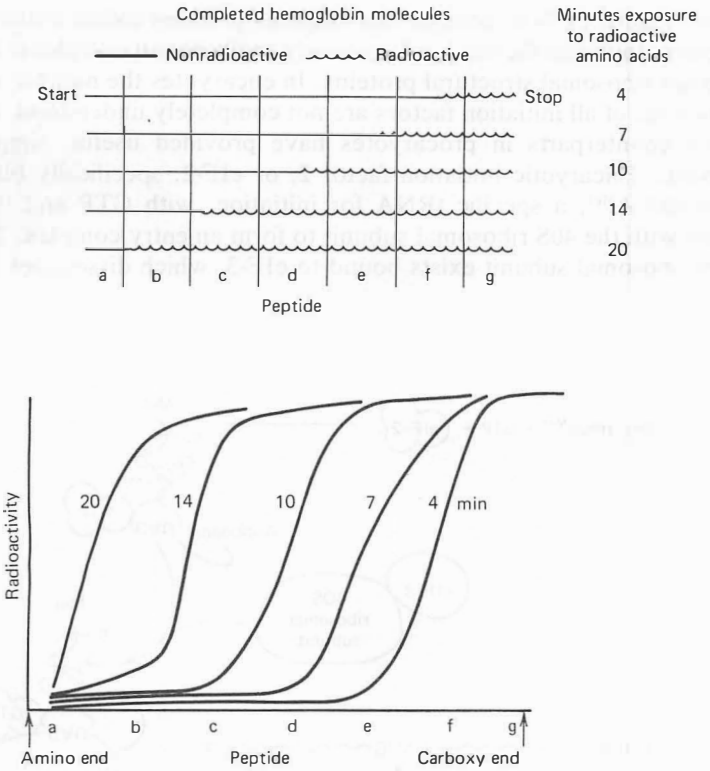


Figure 19.6
Hemoglobin synthesis starts at the amino end and stops at the carboxy end.
See text for explanation.

attach and detach from mRNA at random, that a given ribosome remains bound to a given mRNA molecule until the message is completely translated, and that in the translation process there is movement of each ribosome relative to the mRNA.

Initiation of Protein Synthesis

Three operational divisions of protein synthesis provide a useful descriptive framework. They are initiation, elongation, and termination. Initiation consists of the placement of the amino-terminal amino acid and involves, first, the formation of a complex on the 40S ribosomal subunit and, second, the addition of the 60S ribosomal subunit to give the whole ribosome 80S initiation complex. The pro-

cess (Figure 19.7) is intricate and requires proteins called initiation factors. Initiation factors bind reversibly to ribosomal complexes but are not ribosomal structural proteins. In eucaryotes the number and functions of all initiation factors are not completely understood, but their counterparts in procaryotes have provided useful, simpler models. Eucaryotic initiation factor 2, or eIF-2, specifically binds Met-tRNA$_i^{Met}$, a specific tRNA for initiation, with GTP and then joins with the 40S ribosomal subunit to form an entry complex. The 40S ribosomal subunit exists bound to eIF-3, which dissociates the

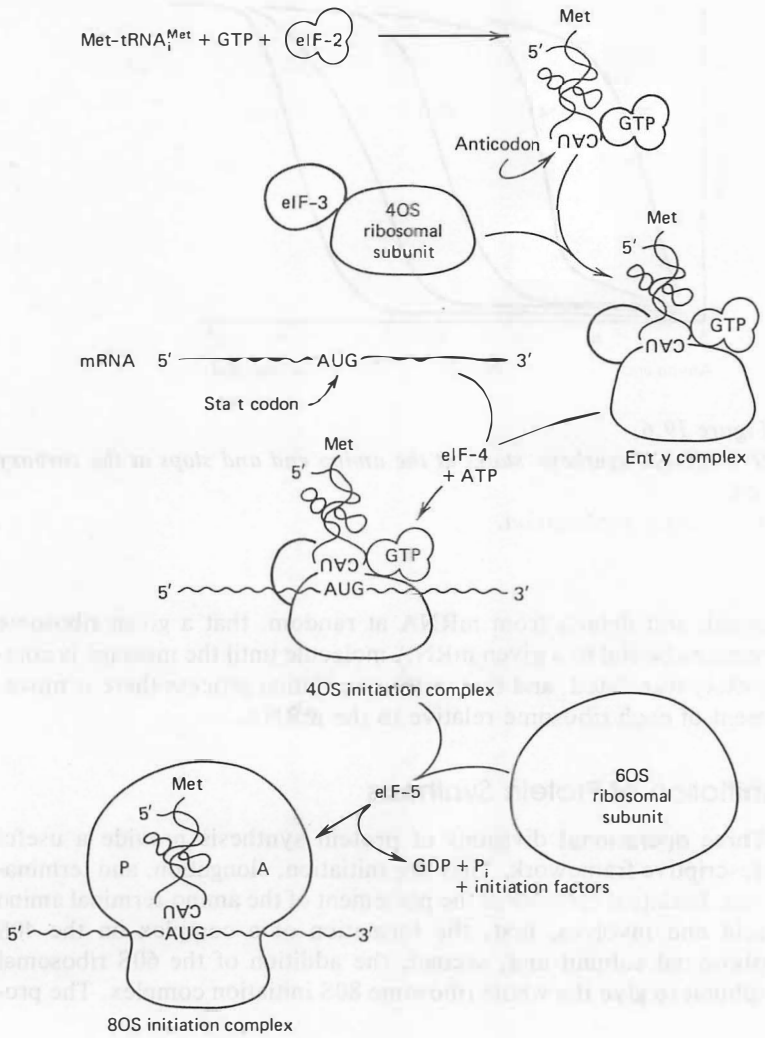


Figure 19.7
A diagram of initiation of protein synthesis. *See text for description. Shown is initiation for the first ribosome. Later entry complexes add not to "naked" mRNA but to a polyribosome.*

40S and 60S ribosomal subunits. That dissociation is required in each cycle of protein synthesis as done by each ribosome. Whereas its procaryotic counterpart is a single polypeptide chain, eIF-3 has 9–11 subunits, testifying to the greater complexity and/or to our lesser understanding of eucaryotic initiation.

Met-tRNA$_i^{Met}$ functions only in initiation. Conversely, Met-tRNA$_m^{Met}$ (note the subscript) is required for elongation and does not place methionine residues in initiator (amino-terminal) position. Met-tRNA$_m^{Met}$ is not recognized by eIF-2.

One mechanism of control of protein synthesis involves inactivation of eIF-2 by phosphorylation of its α subunit, its larger β and γ subunits being unaffected. The kinase involved is controlled by heme, and we shall return to this control in a subsequent discussion of balance in hemoglobin synthesis. Phosphorylation of the α subunit of eIF-2 is also mediated by interferon. Interferon is a protein produced in cells infected with viruses. We know that interferon stops both transcription and translation (protein synthesis), but the precise sequence of events stemming from its presence is unknown. One action of interferon is to activate a protein kinase, which phosphorylates the α subunit of eIF-2 at a site indistinguishable from that phosphorylated by the heme-controlled kinase. The inactive eIF-2 is unable to form Met-tRNA$_i^{Met}$ 40S entry complex, and initiation stops.

The next step in initiation is the joining of the entry complex and mRNA to form the 40S initiation complex, mediated by at least three factors here grouped by the designation eIF-4, with codon–anticodon pairing. ATP is required, but mechanism and stoichiometry remain obscure. The recognition of the start codon, AUG, by the anticodon in Met-tRNA$_i^{Met}$ (Figure 19.1) sets the reading frame.

In the final step of initiation eIF-5 mediates formation of the 80S initiation complex (Figure 19.7) by addition of the 60S subunit to the 40S initiation complex, with concomitant release of initiation factors and required breakdown of GTP to GDP and P_i. In parallel one of the procaryotic initiation factors has a ribosome-dependent GTPase activity.

Other initiation factors have been partially characterized and are found to be stimulatory, but not absolutely required, for various steps of initiation in eucaryotes. Perhaps their roles will be found to relate to control when fully elucidated.

As shown in Figure 19.7, Met-tRNA$_i^{Met}$ is bound at the P site in the 80S initiation complex. The A site is the entry position for new aminoacyl-tRNAs in elongation. The P site bears only the peptidyl-tRNA during elongation and holds the nascent chains in readiness for addition of the next amino acid residue. The two sites are experimentally differentiated and defined by sensitivity to attack by the antibiotic, puromycin, which is discussed below.

Elongation of Peptide Chain

Elongation is the second and major division of protein biosynthesis. Elongation consists of placement of all of the amino acid residues except the first one and involves three basic events, repeated for each residue: (1) binding of aminoacyl-tRNA at the A site; (2) formation of the peptide bond; and (3) movement of the new peptidyl-tRNA to the P site.

The aminoacyl-tRNA specified by the next codon in the initiation complex does not simply join to that complex. Rather, this incoming tRNA forms an entry complex with GTP and a protein called elongation factor 1. This entry complex relinquishes its aminoacyl-tRNA, as specified by the next codon, to the A site on the 80S initiation complex. With this binding the GTP is released as GDP and P_i. Elongation factor 1 forms complexes with aminoacyl-tRNAs for all amino acids, including Met-$tRNA_m^{Met}$, but does not bind Met-$tRNA_i^{Met}$, whose carriage to the ribosome is mediated only by initiation factor 2. With completion of this binding step, Met-$tRNA_i^{Met}$ and the next aminoacyl-tRNA now lie juxtaposed on the ribosome, as shown in Figure 19.8, but no peptide bond exists yet.

Peptidyltransferase, an enzyme constituting one of the proteins of the large ribosomal subunit of procaryotes and believed to occupy a similar, strategic surface location in the 60S subunit of eukaryote ribosomes, now catalyzes a nucleophilic attack by the free amino group of the second aminoacyl-tRNA on the carbonyl carbon of Met-$tRNA_i^{Met}$, to form the first peptide bond, as shown in Figure 19.9. With completion of this step, $tRNA_i^{Met}$ remains bound at the P site, ready to be discarded, and dipeptidyl-tRNA is bound at the A site, ready to move to the P site in the next step. Formation of the peptide bond requires no energy source other than the aminoacyl-tRNA here represented by Met-$tRNA_i^{Met}$.

The final event of elongation is translocation. Dipeptidyl-tRNA in codon–anticodon register with the codon for the second amino acid (leucine in Figure 19.9) now moves to the P site under mediation of translocase, also called elongation factor 2, and $tRNA_i^{Met}$ is released simultaneously. For its function EF-2 requires GTP hydrolysis to GDP and P_i, and EF-2 has a ribosome-dependent GTPase activity. Presentation of translocation as a separate event may be misleading because translocation occurs simultaneously with the first event for the next amino acid, that is, the binding of aminoacyl-tRNA. Thus, as translocation occurs, elongation factor 1 places the third aminoacyl-tRNA at the A site, as described for the second aminoacyl-tRNA. In the combined event the precise stoichiometry of GTP hydrolysis remains obscure. The large subunit of the procaryotic ribosome contains two copies of a GTPase called the A

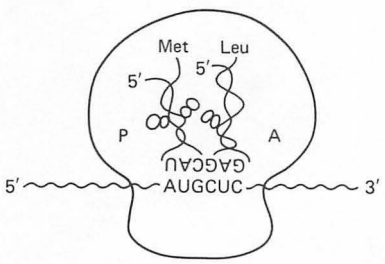


Figure 19.8
The first step of elongation is placement of the second aminoacyl-tRNA, here Leu-tRNA, in the A site, the position for peptide bond formation.
See text for description.

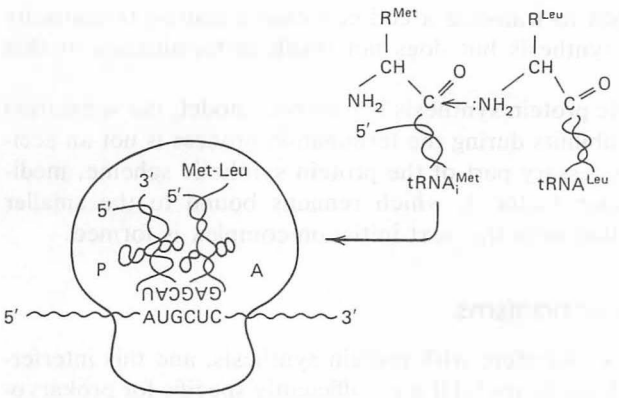


Figure 19.9
Peptidyltransferase catalyzes the second step of elongation with formation of the peptide bond.

protein, which may be involved in these events. Completion of the translocation-binding step results in the complex shown in Figure 19.10. Repetition of the peptidyltransferase step and of the translocation-binding step places all remaining amino acid residues in the nascent protein.

Termination of Protein Synthesis

Termination is the final operational division of protein biosynthesis. Termination occurs after placement of the final amino acid residue at the carboxy terminal of the nascent protein and consists of release of the newly synthesized protein, release of the tRNA of the final amino acid residue, and release of mRNA and ribosomal subunits, all in response to a termination signal. The termination signal is one or more of the codons UAG, UAA, or UGA. These termination signals are recognized not by a tRNA, but by a protein release factor, which requires GTP to bind to ribosomes and has a ribosome-dependent GTPase activity that is stimulated by termination codons. In the release reaction GTP is split to GDP and P_i. Peptidyltransferase activity is required for release to occur. The release factor allows water to substitute for the α-amino group of an amino acid in the nucleophilic attack (Figure 19.9) on the carbonyl carbon of the terminal amino acid residue. The termination codon is a specific requirement, and it does not simply represent a codon for which there is no tRNA. Thus the simple absence of the required

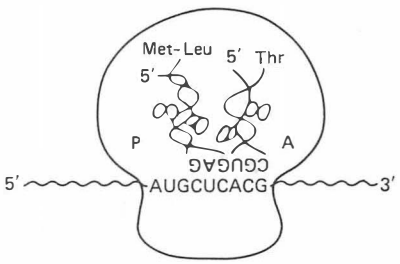


Figure 19.10
After translocation and simultaneous binding of the third aminoacyl-tRNA, here Thr-tRNA, at the A site, Met-Leu-tRNA is in the P site. *Peptide bond formation will follow.*

aminoacyl-tRNA to translate a codon within a cistron temporarily stops protein synthesis but does not result in termination at that point.

If procaryotic protein synthesis is a correct model, the separation of ribosomal subunits during the termination process is not an accidental but a necessary part of the protein synthetic scheme, mediated by initiation factor 3, which remains bound to the smaller ribosomal subunit until the next initiation complex is formed.

Antibiotic Mechanisms

Many antibiotics interfere with protein synthesis, and this interference may be clinically useful if it is sufficiently specific for prokaryotic vis-à-vis eucaryotic protein synthesis. Some of the clinically useful antibiotics that act by inhibition of protein synthesis are the tetracyclines, the aminoglycosides, chloramphenicol, and erythromycin. Tetracyclines work by binding to the procaryotic ribosome and preventing the normal binding of aminoacyl-tRNA at the A site. Streptomycin and other aminoglycosides work by binding to the small (30S) subunit of the procaryotic ribosome, distorting it, and thereby producing incorrect codon–anticodon interaction with consequent misreading of the genetic code. A particular ribosomal protein, S12, constitutes the binding site for streptomycin, and organisms with mutant forms of this protein may be resistant to or even dependent upon streptomycin. Erythromycin interferes with prokaryotic protein synthesis by inhibiting translocase. Chloramphenicol specifically inhibits procaryotic peptidyltransferase. The antibiotic also inhibits mitochondrial ribosomal peptidyltransferase, this inhibition being one measure of similarity between mitochondria and procaryotes. Cycloheximide inhibits eucaryotic but not prokaryotic peptidyltransferase, and thus is the converse of chloramphenicol in its activity. Of course, the antibiotic has no clinical usefulness for this reason.

Many antibiotics too toxic for clinical use are important in experimental work and have sometimes played an important role in elucidating mechanisms of various steps of protein synthesis. Puromycin is perhaps the best example because reactivity with puromycin defines whether a peptidyl-tRNA resides at the P site or at the A site of a ribosome. Puromycin has a structure closely resembling the 3′ end of aminoacyl-tRNA (Figure 19.11) and stops protein synthesis instantly by completing a peptide bond, under mediation of peptidyltransferase, with the nascent peptidyl-tRNA on the P site. The peptide, now with carboxy-terminal puromycin in covalent linkage, is immediately released from the ribosome, having no tRNA to hold it to the A site. Puromycin does not react with peptidyl-tRNA bound at the A site.

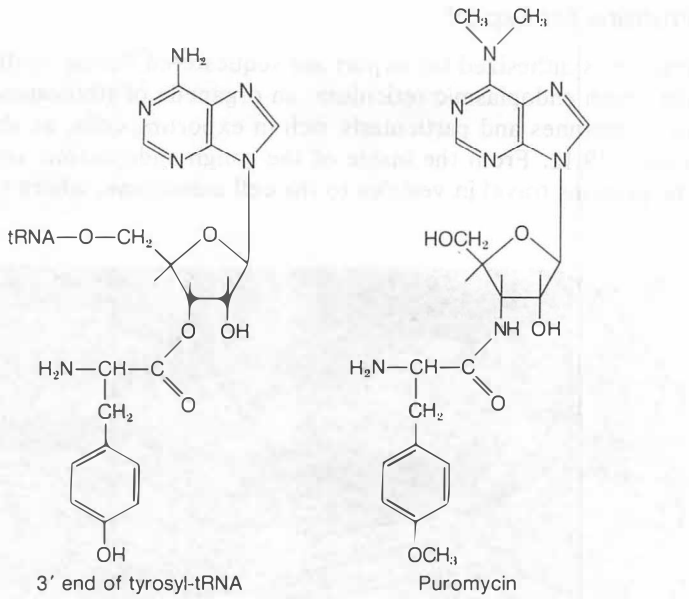


Figure 19.11
Puromycin (right) interferes with protein synthesis by functioning as an analog of aminoacyl-tRNA, here tyrosyl-tRNA (left) in the peptidyltransferase reaction.

Another clinically relevant example of inhibition of protein synthesis is the inactivation of mammalian translocase, elongation factor 2, by the toxins of *Corynebacterium diphtheriae* and *Pseudomonas aeruginosa.* These protein toxins bind to the cell membrane, then enter the cell either intact (*Pseudomonas* toxin) or as an active subunit (diphtheria toxin). Once inside the cell the protein functions as an enzyme, catalyzing the reaction

$$\text{Translocase} + \text{NAD}^+ \xrightarrow{\text{toxin}} \text{ADPR-translocase} + \text{nicotinamide} + \text{H}^+$$

The translocase, thus covalently modified, is irreversibly inactivated under in vivo conditions. Thus the extraordinary potency of these toxins results from their catalytic rather than stoichiometric mode of action. Perhaps one molecule of toxin is sufficient to kill a cell. Consistent with this mode of action, these toxins kill cells slowly in contrast to other bacterial toxins.

Proteins for Export

Proteins synthesized for export are sequestered during synthesis in the rough endoplasmic reticulum, an organelle of ribosomes bound to membranes and particularly rich in exporting cells, as shown in Figure 19.12. From the inside of the rough endoplasmic reticulum the proteins travel in vesicles to the cell membrane, where they are

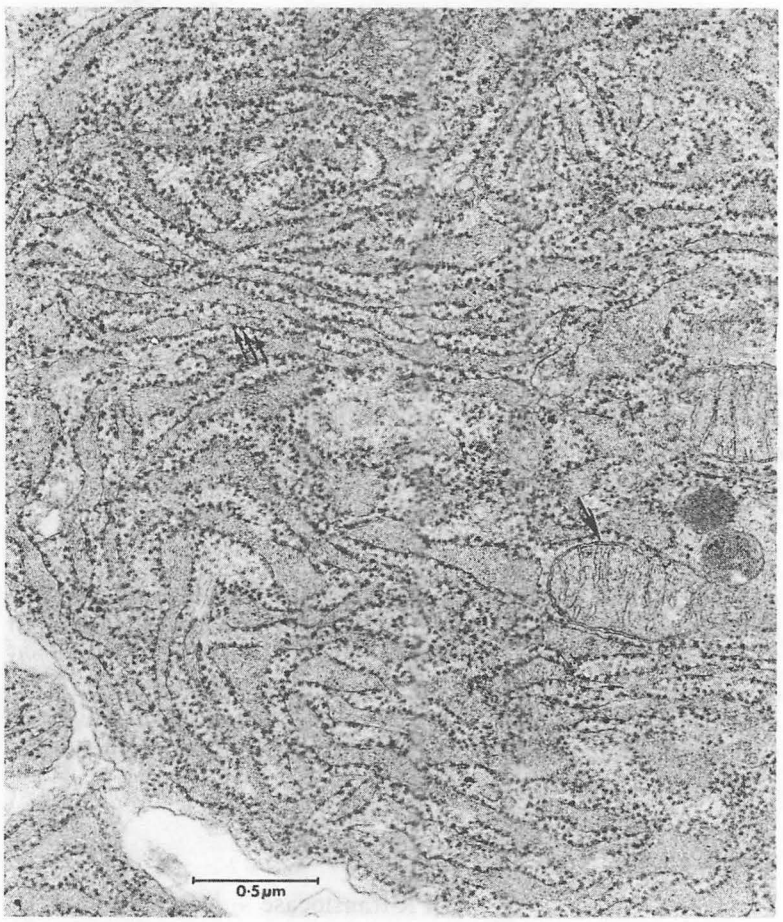


Figure 19.12
Rough endoplasmic reticulum of a plasma cell.
The three parallel arrows indicate three ribosomes among the many attached to the extensive membranes. The single arrow indicates a mitochondrion for comparison.
Courtesy of Dr. U. Jarlfors, University of Miami.

excluded after fusion of the vesicle and cell membranes. The membrane-bound ribosomes of the rough endoplasmic reticulum are identical to and in equilibrium with free ribosomes. They bind by the 60S subunit to the membrane of the rough endoplasmic reticulum through the mediation of the transmembrane glycoproteins, ribophorin I and ribophorin II. However, the binding does not take place until synthesis of the protein for export has begun. At the amino terminal of such proteins are signal peptides, which identify proteins to be exported. The signal peptides are variable but always hydrophobic sequences ranging in length from 15 to 30 amino acid residues. There are a wide variety of different signal peptides, even different immunoglobulin G light chains having different signal peptides. Proteins not to be exported, for example, the α- and β-globins of hemoglobin, do not have signal peptides. As the export protein is synthesized, it is threaded through the rough endoplasmic reticulum membrane by a mechanism not fully understood but thought to involve the action of signal peptidase, an enzyme that cleaves the signal peptide from the nascent export proteins before synthesis of the protein is complete. Curiously, one prominent export protein, ovalbumin, has no signal peptide but does have a hydrophobic sequence at the amino terminal. Ovalbumin competes with other proteins excreted by oviduct cells, following the same pathway. Proteins destined to become membrane proteins also may have transient signal peptides but remain imbedded in the membrane rather than undergoing excretion from the cells for reasons not well understood.

Mitochondrial Protein Synthesis

Synthesis of special proteins for export introduces the compartmentalization of protein synthesis and the involvement of organelles in that compartmentalization. The mitochondria provide an interesting example of protein synthesis occurring entirely within an organelle. Most mitochondrial proteins are synthesized in the cytoplasm on free ribosomes as directed by information in mRNA transcribed from nuclear DNA. The proteins then cross over the double membrane into the mitochondria in a process accompanied by protein cleavage and resembling the entry of diphtheria toxin into cells. However, mitochondria have DNA, which codes not only for rRNA and for tRNAs but for a few mitochondrial proteins. Human mitochondrial DNA is a circular double helix of 16,500 base pairs. Recall that tRNA genes would require less than 100 base pairs and that a protein with molecular weight 37,000 requires a gene with about 1,000 base pairs, not counting leaders and intervening sequences. Interestingly, human mitochondria use the codon UGA, a terminator in all other known systems, as a tryptophan codon in

addition to the single ordinary tryptophan codon, UGG. Moreover, human mitochondria use the codon AUA, an isoleucine codon in all other systems, as a second methionine codon in addition to AUG, the initiator codon. These exceptions to usual codon usage are unique.

The mitochondria have a complete and independent protein synthetic apparatus, including RNA polymerase, ribosomes, aminoacyl-tRNA synthetases, tRNAs, and all other necessary factors. However, the tRNAs, the aminoacyl-tRNA synthetases, and the ribosomes are not the same as those in the cytoplasm. The ribosomes are smaller than cytoplasmic ribosomes and resemble bacterial ribosomes in that respect and in their sensitivity to inhibition of peptidyltransferase by chloramphenicol. In fact the mitochondrial protein synthetic system resembles that of bacteria in other respects. For example, as in procaryotic protein synthesis the initiator in mitochondrial protein synthesis is not Met-$tRNA_i^{Met}$, but formyl-Met-$tRNA_i^{Met}$. Recall that two different tRNAs for methionine, $tRNA_i^{Met}$ and $tRNA_m^{Met}$, are both substrates for methionyl-tRNA synthetase, giving Met-$tRNA_i^{Met}$ and Met-$tRNA_m^{Met}$. Further recall that Met-$tRNA_i^{Met}$ is recognized by initiation factor 2 but not by elongation factor 1, whereas the converse is true for Met-$tRNA_m^{Met}$. A third enzyme that recognizes the difference between $tRNA_i^{Met}$ and $tRNA_m^{Met}$ is transformylase, which catalyzes the formylation of the α-amino group of methionine in Met-$tRNA_i^{Met}$, but not in Met-$tRNA_m^{Met}$, as shown in the reaction

$$N^{10}\text{-Formyltetrahydrofolate} + \text{Met-}tRNA_i^{Met} \longrightarrow \text{tetrahydrofolate} + \text{HCO-Met-}tRNA_i^{Met}$$

Usually the formyl group or the formylmethionine residue is removed from the finished protein. Nevertheless, this is the first mention of a covalently modified amino acid in a protein, and it serves as an introduction to a discussion of posttranslational modification of polypeptides.

19.4 POSTTRANSLATIONAL MODIFICATION OF POLYPEPTIDES

The newly synthesized protein as released from the ribosome may not be in final, functional state. To achieve function posttransla-

tional modifications may be required. Proteins undergo a variety of modifications to provide final structures composed of subunits, final structures containing prosthetic groups, or final structures containing covalently modified amino acid residues.

Consider first the formation of subunit proteins, which may result either from aggregation of subunits or from cleavage of a larger precursor. Hemoglobin formation is an example of the former process. α- and β-Globins, synthesized separately with information from mRNAs transcribed from DNA in different chromosomes, associate with heme and with each other to form hemoglobin. The synthesis of α-globin, β-globin, and heme must be coordinated to produce stoichiometric amounts. The coordination requires control of protein synthesis, a control already implicit in the marked variation between cells of different types in terms of amounts and types of proteins synthesized. Much of the necessary control is exerted at the transcriptional level and is expressed by differing amounts of mRNAs in different cells or in the same cell at different times. If we regard this as gross control, fine tuning may occur in translation. An example of the complexities of such translational control is seen in reticulocytes synthesizing hemoglobin. Here in the presence of constant concentrations of mRNAs for α- and for β-globin, their balanced synthesis is dependent upon presence of heme. In its absence a protein kinase, which is not dependent on cyclic nucleotides, phosphorylates the α-subunit of eIF-2 (Figure 19.13), rendering it unable to form the 40S initiation complex. This protein kinase uses ATP and is called hemin-controlled translation repressor. In addition to decreasing globin synthesis, it decreases α-globin synthesis more than β-globin synthesis, a phenomenon suggesting different affinities of the respective mRNAs for the initiation complex.

Defects in coordination of heme, α-globin, and β-globin synthesis are manifest at the clinical level. For example, iron deficiency decreases heme synthesis and results in hemoglobin deficiency in red blood cells. Thalassemia has already been mentioned as comprising a class of disorders characterized by decreased or absent α- or β-globin with relative excess of the other globin, a state causing

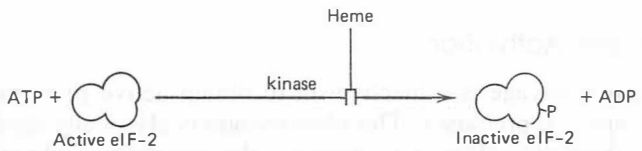


Figure 19.13
Role of heme in blockade of eIF-2 inactivation by protein kinase, termed hemin-controlled translation repressor.

CLIN. CORR. 19.4 FAMILIAL HYPERPROINSULINEMIA

Familial hyperproinsulinemia, an autosomal dominant condition, results from lack of one protease or, more probably, from a structural change in proinsulin, preventing cleavage. Approximately equal amounts of insulin and proinsulin are released into the circulation. Although affected individuals have high levels of proinsulin in the blood, they are apparently normal in terms of glucose metabolism, being neither diabetic nor hypoglycemic.

anemia from decreased red blood cell production and increased red blood cell destruction. Any of the many possible reasons for decreased or absent globins, from a gene deletion (DNA level) to a terminator mutation (page 922) will give the same end result, and thalassemia, with many causative defects, remains the best example of genetic heterogeneity understood at a molecular level.

Biosynthesis of Insulin

Perhaps the simplest mechanism for synthesis of stoichiometric quantities of protein subunits is that using a larger precursor, which is specifically cleaved to provide the subunits or peptide chains. Thus in insulin biosynthesis translation of the mRNA produces preproinsulin, in which "pre-" signifies the signal peptide and "pro-" signifies a precursor of insulin containing both the A and B peptides and the intervening C-peptide, as shown in Figure 19.14. Specific proteases within the endoplasmic reticulum or Golgi apparatus or in the associated vesicles cleave proinsulin to insulin and C-peptide, which are released into the bloodstream with a small amount of uncleaved proinsulin. In familial hyperproinsulinemia an interesting and different situation exists (Clin. Corr. 19.4).

Perhaps the importance of proinsulin synthesis resides in the requirement for proinsulin in correct disulfide bond formation. Unlike the subunits of hemoglobin, which cohere by noncovalent bonding alone, the A- and B-peptides of insulin are held together by disulfide bonds, and the A-peptide contains an intrachain disulfide bond. Although formation of disulfide bonds in proteins may occur spontaneously, correct cysteine-residue specificity requires the correct tertiary structure as dictated by the primary structure, according to the thermodynamic hypothesis. Thus when insulin itself is reduced and denatured, then gently renatured with concomitant reoxidation of thiol groups to form disulfide bonds, those bonds form in random fashion, and native insulin structure and function is not achieved. In contrast, when the same experiment is done with proinsulin, the intact total primary structure assures the correct reformation of native proinsulin, which can be subsequently cleaved to give native insulin.

Zymogen Activation

Precursor cleavage is a mechanism to obtain active proteins from totally inactive precursor. The phenomenon is classically illustrated by zymogen activation, for example, the cleavage of chymotrypsinogen to give chymotrypsin and two inactive peptides (Chapter 24). Inappropriate activation of zymogens in the pancreas for any reason leads to autodigestion expressed clinically as acute pan-

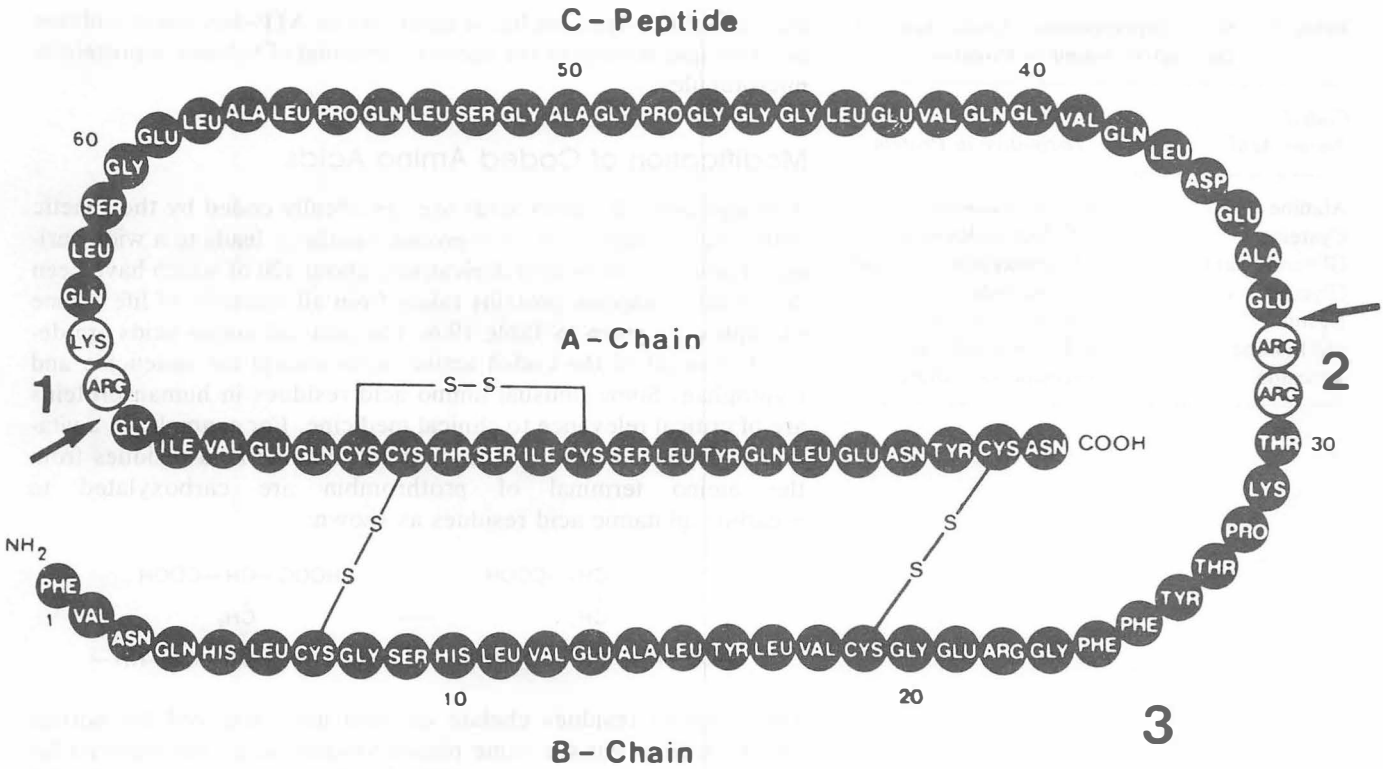


Figure 19.14
Human proinsulin.
After cleavage at the two sites indicated by arrows the arginine residues 31, 32, and 65 and the lysine residue 64 are removed to give insulin and C-peptide.
Reprinted, by permission of the New Eng. J. of Med. 302: 165, 1980.

creatitis. Not all zymogens yield subunit enzymes after cleavage. For example, pepsinogen and trypsinogen are activated by cleavage and removal of amino terminal peptides to leave single-chain, active proteins.

Many proteins are modified by removal or addition of terminal amino acid residues. All proteins do not have an amino-terminal methionine residue, and proteins lacking this initial residue have had it and perhaps further residues removed. Other proteins are modified by addition of amino acid residues. In some cases such addition is mediated by tRNA. For example, arginyl-tRNA is a donor of the arginine residue for the amino terminal of some proteins. In at least one case, however, the addition is not tRNA-dependent. Thus, the

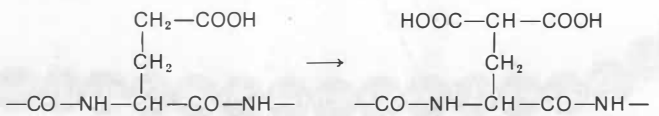
Table 19.6 Some Representative Amino Acid Derivatives Found in Proteins

Coded Amino Acid	*Derivative in Protein*
Alanine	*N*-Acetylalanine
Cysteine	*S*-Galactosylcysteine
Glutamic acid	γ-Carboxyglutamic acid
Glycine	Glycinamide
Lysine	ε-*N*-Methyllysine
Methionine	*N*-Formylmethionine
Tyrosine	Tyrosine O^4-sulfate

enzyme tubulin-tyrosine ligase catalyzes an ATP-dependent addition of a tyrosine residue to the carboxy terminal of tubulin, a protein in microtubules.

Modification of Coded Amino Acids

Although only 20 amino acids are specifically coded by the genetic code, their modification after protein synthesis leads to a wide variety of unusual amino acid derivatives, about 120 of which have been described in various proteins taken from all varieties of life. Some examples are given in Table 19.6. The unusual amino acids are derived from all of the coded amino acids except for isoleucine and tryptophan. Some unusual amino acid residues in human proteins are of critical relevance to clinical medicine. For example, in a vitamin K-dependent reaction the first 10 glutamic acid residues from the amino terminal of prothrombin are carboxylated to γ-carboxyglutamic acid residues as shown:



The modified residues chelate calcium ions, required for normal blood clotting. For the same reason vitamin K is also required for synthesis of clotting factors VII, IX, X. In patients treated with coumarin anticoagulants, which are vitamin K antagonists, an inactive precursor of prothrombin circulates in the plasma, and the related delay in blood clotting is measured by the test known as the prothrombin time. γ-Carboxyglutamic acid residues are also found in protein of cortical bone, where they presumably play a role in binding hydroxyapatite to the organic bone matrix, again through chelation of calcium ions. These comments on covalent modification of amino acid residues serve as an introduction to a more detailed look at extensive modifications required for the biosynthesis of collagen.

19.5 BIOSYNTHESIS OF COLLAGEN

For several reasons the biosynthesis of collagen deserves special attention. Collagen, forming the fibrous network that holds organs

and tissues intact, is the most abundant protein in the human body. Its biosynthesis serves as the most extensive example in protein synthesis of a sequence of multiple, posttranslational, modification steps and of export. Unusual and important features include extracellular processing and spontaneous self-assembly of collagen molecules into fibrils. Various hereditary defects related to collagen biosynthesis cause a number of illustrative diseases and show the importance of mutation in posttranslational modification as distinct from mutations in primary structure.

A brief review of collagen structure (Figure 19.15) is essential for an understanding of its biosynthesis. Briefly, a collagen molecule consists of three α chains, each in a left-handed helix of approximately 1,000 residues, twisted about each other in a right-handed helix with glycine residues at crossing points, occurring every third residue. Thus the approximate formula is $(X—Y—Gly)_{333}$ for a collagen molecule of ~300 nm in length. The collagen molecules aggregate in a parallel, staggered fashion to produce fibrils, which range 10–200 nm in diameter depending on the collagen type. The fibrils are visible in the extracellular matrix of connective tissue by electron microscopy, which reveals a banded structure resulting from the staggered nature of collagen molecules in the fibril, each molecule being displaced from its neighbors by 0, 1, 2, 3, or 4 axial stagger lengths of 234 ± 1 residues. Approximately 100 of the X residues are proline, and approximately 100 of the Y residues are 4-hydroxyproline. The cyclic nature of these residues limits rotation and thereby provides stability. Other amino acids in clusters of hydrophobic and charged residues occupy the other positions designated by X and Y. The different primary structures of the α chains are specified by different genes and give rise to at least five different types of collagen, having different organ distribution. Disorders of collagen structure are listed in Table 19.7 (Clin. Corrs. 19.5–19.9).

In addition to hydroxyproline, finished collagen contains hydroxylysine and glycosylated hydroxylysine. The existence of modifications in coded amino acid residues signals enzymatic mechanisms for their biosynthesis and opens possibilities for defects in those mechanisms. Such defects would be expected to cause disease, just as do direct mutations in primary structure.

Preprocollagen to Procollagen

A consideration of the steps in collagen biosynthesis (Figure 19.16) is necessary for understanding of other disorders. The initial translation product is a procollagen chain with an amino-terminal propeptide of about 180 amino acid residues and a carboxy-terminal propeptide of about 300 amino acid residues destined not to appear in the

CLIN. CORR. 19.5 THE EHLERS–DANLOS SYNDROME

The Ehlers–Danlos syndrome constitutes a group of clinical disorders with various manifestations of structural laxity or defects in connective tissue, all resulting from defects in the structure of collagen. For example, type III collagen is composed of three α_1 III chains and is particularly important in skin, arteries, and the uterus. Type IV Ehlers–Danlos syndrome is caused by a decrease in the amount of type III collagen present in these organs and tissues. Manifestations may be severe, with arterial rupture, intestinal perforation, rupture of the uterus during pregnancy or labor, and easy bruisability of thin, translucent skin. Conversely, type I collagen consists of two α_1 I chains and one α_2 chain. It is the most prevalent form of collagen in adults and is particularly concentrated in bone. The amount of type I collagen present in bone is decreased in osteogenesis imperfecta, a hereditary disorder of bone matrix leading to multiple fractures with resultant severe bony deformities. Decrease in the amount of synthesis of a normal collagen structure implies a defect or defects in control of protein synthesis, but the precise mechanism of the difficulty is not understood for either type IV Ehlers–Danlos syndrome or for osteogenesis imperfecta.

CLIN. CORR. 19.6 SCURVY AND HYDROXYPROLINE SYNTHESIS

In scurvy the severe ascorbic acid deficiency causes decreased hydroxyproline synthesis. Collagen containing insufficient hydroxyproline loses temperature stability and is less stable than normal collagen at body temperature. The resultant clinical manifestations are distinctive and understandable: suppression of the orderly growth process of bone in children, poor wound healing, and increased capillary fragility with resultant hemorrhage, particularly in the skin. Severe ascorbic acid deficiency leads secondarily to a decreased rate of procollagen synthesis.

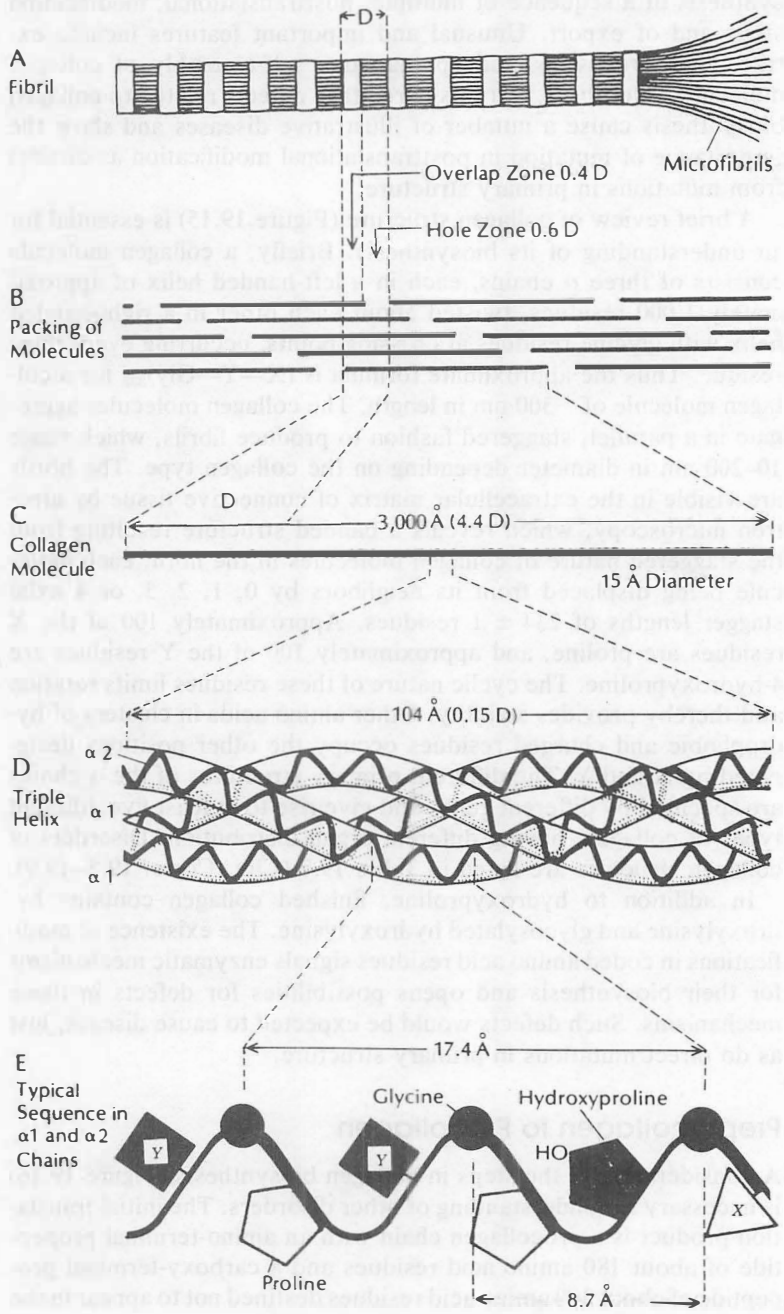


Figure 19.15
Collagen structure, illustrating the regularity of primary sequence, the left-handed α helix, the right-handed triple helix, the 300-nm molecule, and the organization of molecules in a typical fibril, within which the collagen molecules are cross-linked.
Figure by B. Tagawa. Reproduced with permission from D. J. Prockop and N. A. Guzman, Hosp. Pract. 12:61, 1977.

Table 19.7 Some Disorders of Collagen Structure

Disorder	*Collagen Defect*	*Clinical Manifestations*
Ehlers–Danlos IV	Decrease in type III	Arterial, intestinal, or uterine rupture; thin, easily bruised skin
Osteogenesis imperfecta	Decrease in type I	Blue sclerae, multiple fractures, and bone deformities
Scurvy	Decreased hydroxyproline	Poor wound healing, deficient growth; increased capillary fragility
Ehlers–Danlos VI	Decreased hydroxylysine	Hyperextensible skin and joints, poor wound healing, musculoskeletal deformities
Ehlers–Danlos VII	Amino terminal propeptide present	Hyperextensible, easily bruised skin; hip dislocations
Ehlers–Danlos V	Decreased cross-linking	Skin and joint hyperextensibility

final collagen molecule. Actually, a preprocollagen containing signal sequences directs the transfer of the procollagen into the rough endoplasmic reticulum and is removed by signal peptidase as previously discussed. The first modification of procollagen then takes place within the rough endoplasmic reticulum. The procollagen contains cysteine residues in both terminal propeptides. Those at the carboxy terminal become involved in interchain disulfide bonds, and those at the amino terminal become involved in intrachain disulfide bonds, which are essential for subsequent triple helix formation. Whether or not the disulfide bond formation is enzymatic is unclear. Interestingly, no cysteine and no cystine residues are present in mature type I collagen.

A second class of modification required for subsequent triple helix formation is hydroxylation of proline and lysine residues. Hydroxyproline is an amino acid found only in collagen. Three different enzymes are required: prolyl-4-hydroxylase, prolyl-3-hydroxylase, and lysylhydroxylase. All require ferrous ion and ascorbic acid, and all use O_2 and α-ketoglutarate in the general reaction

$$\text{AA residue} + O_2 + \alpha\text{-ketoglutarate} \xrightarrow{\text{hydroxylase}} \text{HO-AA residue} + CO_2 + \text{succinate}$$

The amino acid residue must be in a particular sequence in a nonhelical region. Thus lysylhydroxylase acts only on a lysyl residue in the

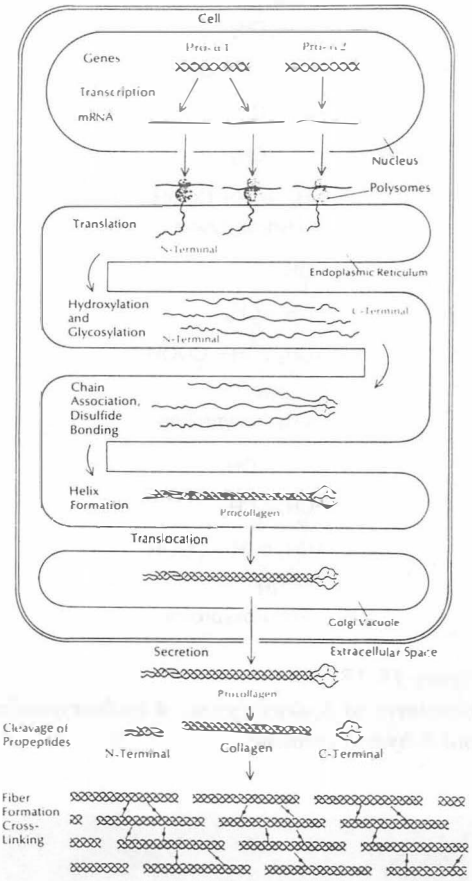
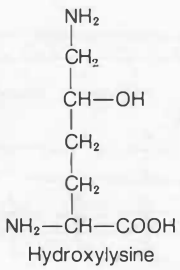
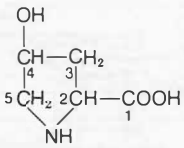


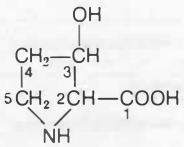
Figure 19.16
Diagram of collagen biosynthesis, showing transcription (only two genes are depicted), translation in the endoplasmic reticulum, various and extensive posttranslational modifications, helix formation, secretion into the extracellular space, cleavage of propeptides, and cross-linking to give fibrils.
Figure by B. Tagawa. Reproduced with permission from D. J. Prockop and N. A. Guzman, Hosp. Pract. 12:61, 1977.



Hydroxylysine



4-Hydroxyproline



3-Hydroxyproline

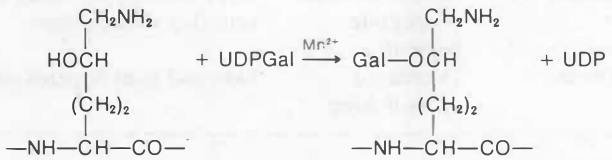
Figure 19.17
Structures of hydroxylysine, 4-hydroxyproline, and 3-hydroxyproline.

CLIN. CORR. 19.7 DEFICIENCY OF LYSYLHYDROXYLASE

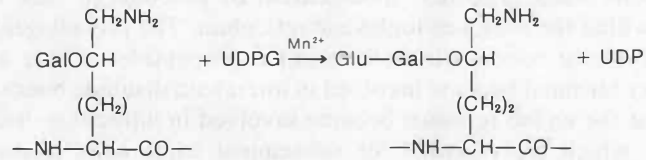
In type VI Ehlers–Danlos syndrome lysylhydroxylase is deficient. As a result collagen with decreased hydroxylysine content is synthesized, and subsequent cross-linking of collagen fibrils, is less stable. The clinical features include marked hyperextensibility of the skin and joints, poor wound healing, and musculoskeletal deformities. Some patients with this form of Ehlers–Danlos syndrome have a mutant form of lysylhydroxylase with a higher

Y position of (X—Y—Gly). Similarly, prolyl-4-hydroxylase acts only on a prolyl residue in the Y position. Prolyl-3-hydroxylase is specific for prolyl residues in the X position but only when a hydroxyprolyl residue is already present in the Y position. The structures of hydroxylysine, 4-hydroxyproline, and 3-hydroxyproline are shown in Figure 19.17.

Further modifications of certain of the hydroxylysyl residues are required for subsequent stable helices and must be performed before the collagen molecule assumes the helical conformation. The modifications are glycosylations and require two enzymes, a galactosyltransferase and a glucosyltransferase. Both enzymes require a divalent cation, preferably Mn^{2+}. The galactosyltransferase uses UDPGal as a donor to add galactosyl residues to specific hydroxylysyl residues,



and the glucosyltransferase uses UDPG as a donor to add glucosyl residues to certain of the newly formed galactosylhydroxylysyl residues.



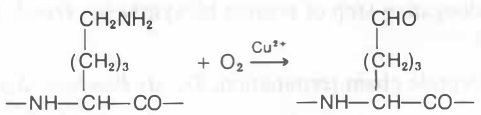
When collagen helices do not form because of any of the defects mentioned in Clin. Corrs. 19.5–19.7, the nonfunctional collagen precursor is secreted slowly by the cells. Such a product is essentially gelatin, the familiar protein that we obtain as food by the heat denaturation of collagen. Normal helical procollagen with disulfide bonds, hydroxylations, and glycosylations now passes through the Golgi complex and is extruded from the cell. Further processing of this helical procollagen product is extracellular.

Procollagen to Collagen

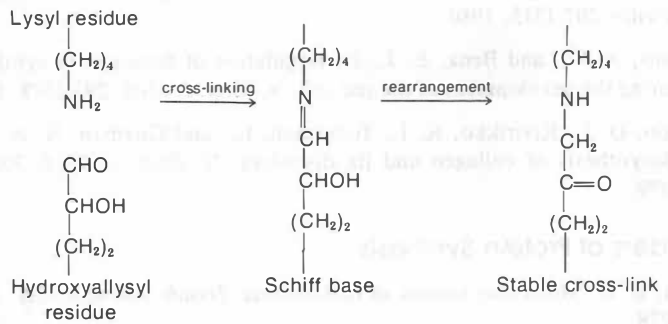
The first steps of extracellular processing involve the removal of the propeptides. Two enzymes are involved: procollagen carboxypeptidase, which removes the carboxy-terminal propeptide, and procollagen aminopeptidase, which removes the amino-terminal propep-

tide. After normal removal of propeptides spontaneous self-assembly of collagen molecules into collagen fibrils ensues. The amino acid sequence alone provides the necessary information for correct fibril formation, and the fibrils formed are remarkably different for the different types of collagen. For example, type I and type II collagen have a totally different appearance by electron microscopy.

Although the collagen at this stage is indistinguishable by electron microscopy from mature collagen, intermolecular cross-linking is required for the necessary tensile strength of mature collagen. The first step and the only enzymatic step in cross-link formation is the oxidative deamination of certain lysyl and hydroxylysyl residues to give the corresponding aldehydes. This reaction is catalyzed by lysyl oxidase, an enzyme requiring cuprous ion and molecular oxygen.



The aldehyde groups undergo two types of linking reactions, probably neither one of which is enzymatic. One type of linking is aldol condensation, mostly between two aldehyde groups on the same α chain. The second type of linkage is Schiff base formation between aldehydes and ε-amino groups of lysyl, hydroxylysyl, and glycosylated hydroxylysyl residues, predominantly in neighboring α chains. The Schiff bases are not stable but serve as precursors to further reactions, leading to stable cross-links. For example, when a hydroxylysyl residue is the source of the aldehyde residue (hydroxyallysyl residue), the compound derived by Schiff base formation with a lysyl residue on another chain can be stabilized by Amadori rearrangement, as shown:



Ultimately, the aldehydes derived from hydroxylysyl residues give the most stable cross-links, a point underscoring the importance of hydroxylysyl formation.

Michaelis constant for ascorbic acid than the normal enzyme. Accordingly, they respond to high doses of ascorbic acid.

CLIN. CORR. 19.8 SKIN FRAGILITY

In Ehlers–Danlos syndrome, type VII, and in dermatosparaxis of cattle and sheep a structural mutation in one procollagen chain interferes with removal of the amino terminal propeptide. The result in cattle and sheep is a fragility of skin so extreme as to be incompatible with life. In man the skin bruises easily and is hyperextensible, but the major manifestation is bilateral hip dislocation.

CLIN. CORR. 19.9 DEFICIENCY OF LYSYL OXIDASE

In type V Ehlers–Danlos syndrome and in some forms of cutis laxa there is a deficiency in lysyl oxidase with consequent cross-linking defects in both collagen and elastin. In cutis laxa the defect is manifested in loose skin, which appears excessively wrinkled, hangs in folds, and lacks the elastic qualities of the hyperextensible skin in Ehlers–Danlos syndrome. In type V Ehlers–Danlos syndrome the skin hyperextensibility is accompanied by joint hypermobility, but the latter is limited to the digits. Copper-deficient animals have deficient cross-linking of elastin and collagen, apparently because of the requirement for cuprous ion by lysyl oxidase. A woman taking high doses of the copper-

chelating drug, D-penicillamine, gave birth to an infant with an acquired Ehlers–Danlos-like syndrome, which subsequently cleared. Side effects of D-penicillamine therapy include poor wound healing and hyperextensible skin.

The disorders of collagen biosynthesis serve well to illustrate the consequences of different types of mutational defects: those affecting regulation of synthesis with formation of inadequate quantities of normal collagen, those with defects in procollagen primary structure, and those characterized by defects in modifying enzymes. These disorders also serve as a rich source of information on the molecular basis for genetic heterogeneity.

BIBLIOGRAPHY

Biosynthesis of Proteins

Clark, B. The elongation step of protein biosynthesis. *Trends Biochem. Sci.* 5:207, 1980.

Caskey, C. T. Peptide chain termination. *Trends Biochem. Sci.* 5:234, 1980.

Hunt, T. The initiation of protein synthesis. *Trends Biochem. Sci.* 5:178, 1980.

Leader, D. P. Protein biosynthesis on membrane-bound ribosomes. *Trends Biochem. Sci.* 4:205, 1979.

Pollack, M. Pseudomonas aeruginosa Exotoxin A. *N. Engl. J. Med.* 302:1360, 1980.

Uy, R., and Wold, F. Post-translational covalent modification of proteins. *Science* 198:890, 1977.

Wickner, W. Assembly of proteins into membranes. *Science* 210:861, 1980.

Special Proteins

Eyre, D. R. Collagen: molecular diversity in the body's protein scaffold. *Science* 207:1315, 1980.

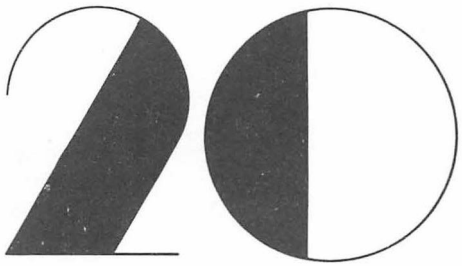
Nienhuis, A. W., and Benz, E. J., Jr. Regulation of hemoglobin synthesis during the development of the red cell. *N. Engl. J. Med.* 297:1318, 1977.

Prockop, D. J., Kivirikko, K. I., Tuderman, L., and Guzman, N. A. The biosynthesis of collagen and its disorders. *N. Engl. J. Med.* 301:13, 1979.

Disorders of Protein Synthesis

Forget, B. G. Molecular lesions in thalassemia. *Trends Biochem. Sci.* 3:86, 1978.

Lagerkvist, U. Codon misreading: a restriction operative in the evolution of the genetic code. *Am. Sci.* 68:192, 1980.



Regulation of Gene Expression

EDWARD GLASSMAN

20.1 OVERVIEW

Cells are highly structured entities in which chemicals are ordered in strict arrangements that lead to complex systems of organized structure and function. The correlation of molecular structure with the function of that molecule constitutes one of the major goals of molecular biology. Current thinking puts a great deal of emphasis on the basic patterns of the flow of chemical information from DNA to RNA to protein and on the mechanisms regulating this flow in order that cells have the types and numbers of proteins and other molecules to meet their needs for structure, function, and maintenance. It makes sense for a cell to vary the amounts of different proteins depending on need; indeed, the bacterium *Escherichia coli* (*E. coli*) makes many more molecules of a protein involved in extensively used pathways than a protein that is little used. As an example, it has been estimated that under some conditions some enzymes that metabolize glucose may be present in excess of 100,000 molecules per cell, while β-galactosidase, an enzyme that metabolizes lactose, is usually represented by about five molecules per cell. Yet the amount of β-galactosidase will increase 1,000-fold if lactose rather than glucose is used as the carbon-energy source by the cell. This difference in amount of various protein molecules is typical of bacterial cells.

The differentiated cells of higher organisms show even greater variation in complexity, chemical composition, physical structure, and biological function. In addition, during mammalian development the appearance and disappearance of specific proteins is well controlled with respect to sequence of events and timing. This overall process is referred to as the control and regulation of *gene expression;* that is, the development of the entire phenotype of the cell and organism. Figure 20.1 presents the many opportunities during the synthesis and processing of RNA and protein wherein regulation of gene expression can, and probably does, take place.

Not surprising is the finding that the underlying mechanisms of the control of information flow vary for different genes, different systems, and different organisms, depending on specific needs for regulation. In this chapter we explore the control of information flow from DNA to RNA; specifically this will involve the regulation of the synthesis of mRNA in bacteria and in mammals. Studies on the control of transcription in bacteria have provided exciting ideas on the process that may prove useful to eventually understanding these processes in mammals. In general, the data suggest that in bacteria gene expression is controlled mainly at the transcriptional level, while mammals also utilize controls at the posttranscriptional

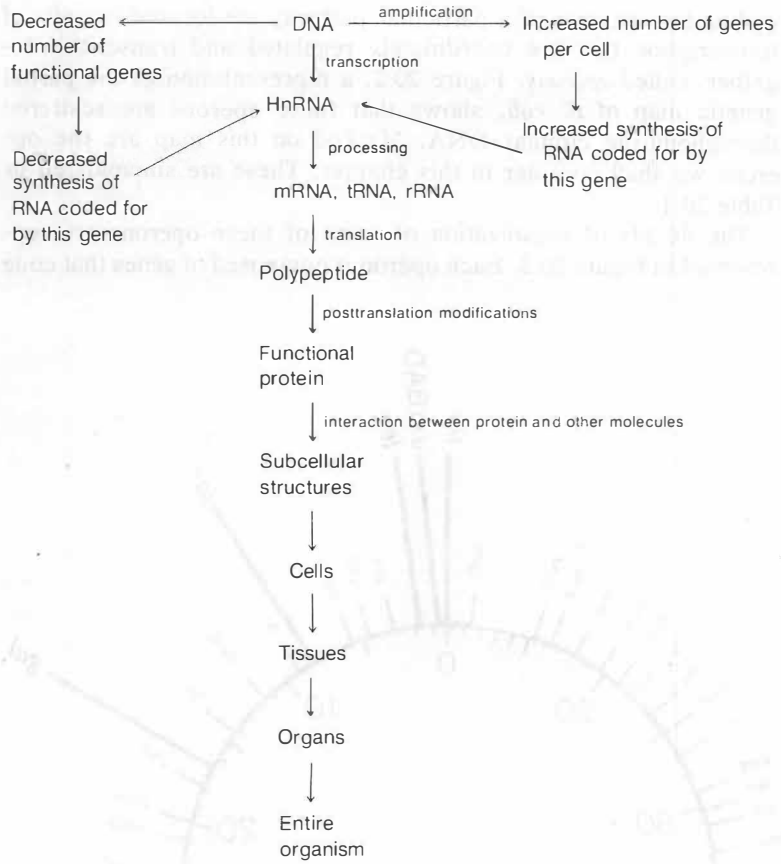


Figure 20.1
Steps in gene expression in mammals at which regulation can occur.
Gene expression in bacteria is simpler.

and translational levels. These two latter processes are discussed in Chapters 18 and 19.

20.2 THE UNIT OF TRANSCRIPTION IN BACTERIA: THE OPERON

The genes coding for the enzymes of bacteria are not located at random on the DNA; instead there is clustering so that the gene's

coding for enzymes of a particular pathway are located in units of transcription that are coordinately regulated and transcribed together, called *operons.* Figure 20.2, a representation of the partial genetic map of *E. coli,* shows that these operons are scattered throughout the circular DNA. Marked on this map are the operons we shall consider in this chapter. These are summarized in Table 20.1.

The details of organization of some of these operons are represented in Figure 20.3. Each operon is composed of genes that code

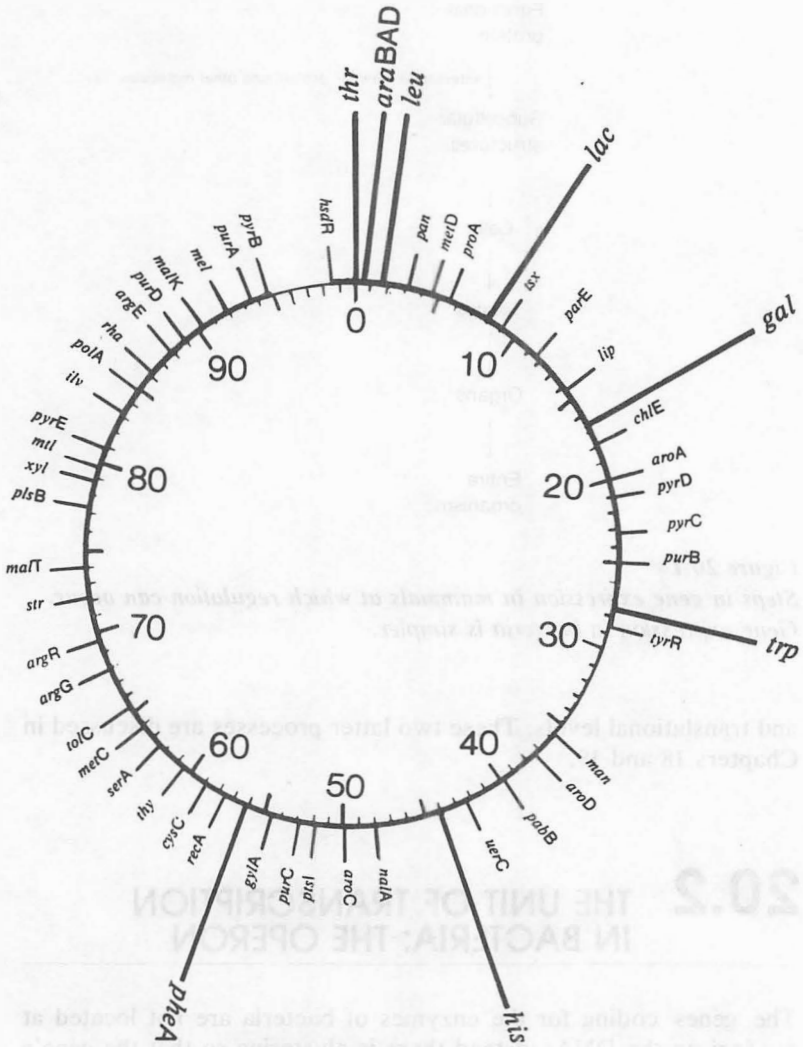


Figure 20.2
***A partial genetic map of* E. coli.**
The locations of only a few of the genes identified and mapped in E. coli *are shown here. The operons discussed in this chapter are marked.*

Reproduced with permission from G. S. Stent and R. Calendar, *Molecular Genetics, An Introductory Narrative,* W. H. Freeman, San Francisco, 1978, p. 289; modified from B. J. Bachmann, K. B. Low, and A. L. Taylor, *Bacteriol. Rev.*, 40:116, 1976.

Table 20.1 Some Operons in *E. coli*

Operon	*Location on Genetic Map (Figure 20.2)*	*Number of Enzymes Coded for*	*Biochemical Function of Metabolic Pathway*
araBAD	1.3	3	Transport and utilization of arabinose as a carbon/energy source
gal	16.5	3	Utilization of galactose as a carbon/energy source; conversion of glucose or galactose to UDP-galactose
his	44.1	10	Synthesis of histidine from 5-phosphoribosyl 1-pyrophosphate
lac	7.8	3	Transport and utilization of lactose as a carbon/energy source
leu	1.9	4	Synthesis of leucine from α-ketoisovalarate
phe	56.1	2	Synthesis of phenylalanine from chorismic acid
thr	0.1	3	Synthesis of threonine from β-aspartic semialdehyde
trp	27.5	4	Synthesis of tryptophan from chorismate

for the amino acid sequence of the enzymes of the pathway; such genes are called *structural genes.* In addition, each operon contains areas of the DNA that are involved with the *coordinate regulation* of the transcription of the structural genes of the operon; these regulatory regions contain a region to which the RNA polymerase attaches, called *promoter regions.* Regulatory regions contain nucleotide sequences called *CRP sites, operator regions,* or *attenuator regions,* depending on the mechanism of transcription regulation. As yet undescribed regulatory regions may exist. Note in Figure 20.3 that different operons have different regulatory regions. Thus the *lac* operon contains a CRP site and an operator region, but no attenuator; the *trp* operon contains an operator region, an attenuator region, a second internal promoter, but no CRP site; the *his* operon contains only an attenuator region. Usually located outside of the operon, and transcribed independently of it, are genes that code for

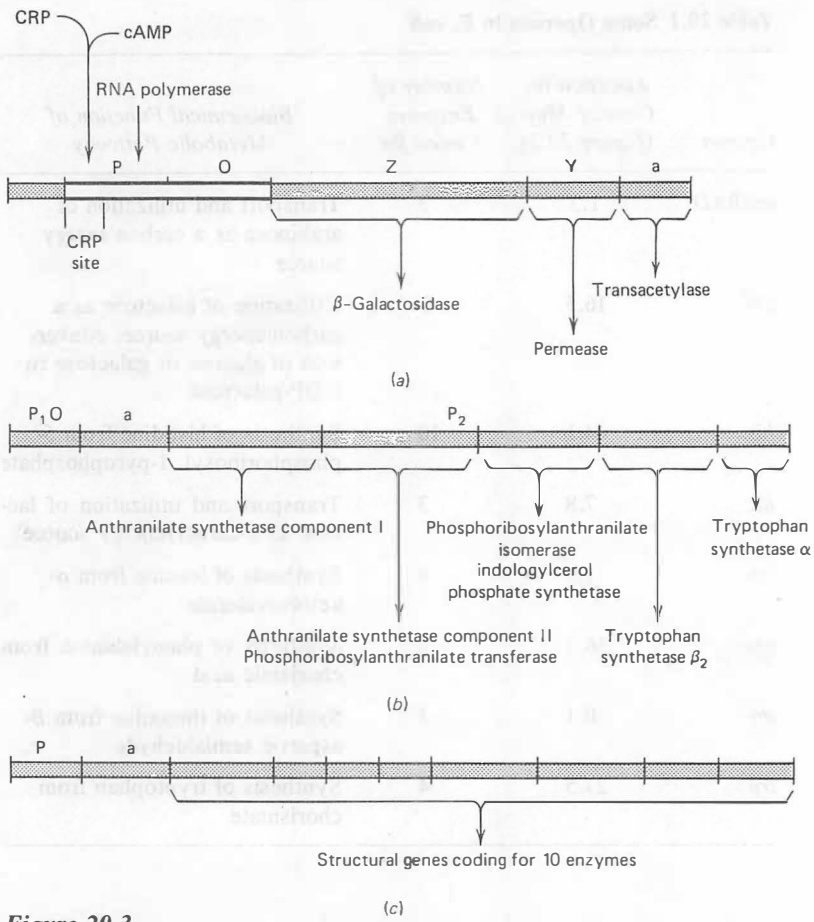


Figure 20.3
The lac, trp, and his operons of bacteria showing differences in regulatory regions.
(a) The lac operon: P = promoter, O = operator. (b) The trp operon: P_1 = main promoter, O = operator, a = attenuator, P_2 = secondary internal promoter. (c) The his operon: P = promoter, a = attenuator.

regulatory proteins that are involved in the regulation of transcription of operons; such genes are called *regulator genes,* even though they serve as structural genes for regulatory proteins. The combination of regulatory proteins with appropriate small molecules of the cell determines whether these proteins will attach to the DNA of the regulatory regions of the operon, and thus affect transcription. Indeed, in many cases, it is the level of the metabolites of the cell that controls

Table 20.2 The Regulation of the Transcription of Some Bacterial Operons

Metabolite	*Operon*	*Regulatory Site Affected*	*Regulatory Protein*	*Other Factors Involved*
Glucose	*lac, gal, araBAD*	CRP site	CRP protein	AMP
Lactose	*lac*	Operator	Repressor	—
Galactose	*gal*	Operator	Repressor	—
Arabinose	*araBAD*	Operator	Repressor	—
Tryptophan	*trp*	Operator	Repressor	—
		Attenuator	—	Ribosomes, tRNA
Threonine	*thr*	Attenuator	—	Ribosomes, tRNA
Histidine	*his*	Attenuator	—	Ribosomes, tRNA
Leucine	*leu*	Attenuator	—	Ribosomes, tRNA
Phenylalanine	*phe*	Attenuator	—	Ribosomes, tRNA
—		Transcription terminator sites	N protein of λ phage	—
Amino acids	rRNA, tRNA	—	—	ppGpp, pppGpp, ribosomes, tRNA

transcription. Table 20.2 lists some of the metabolites that are discussed in this chapter, the operons with which each interacts, the regulatory site and protein involved, as well as other factors. This chapter focuses on elucidating the material presented in Table 20.2.

20.3 THE PROMOTER REGION: AFFINITY FOR RNA POLYMERASE

RNA polymerase does not attach to the bacterial DNA at random; instead there appear to be specific nucleotide sequences, called

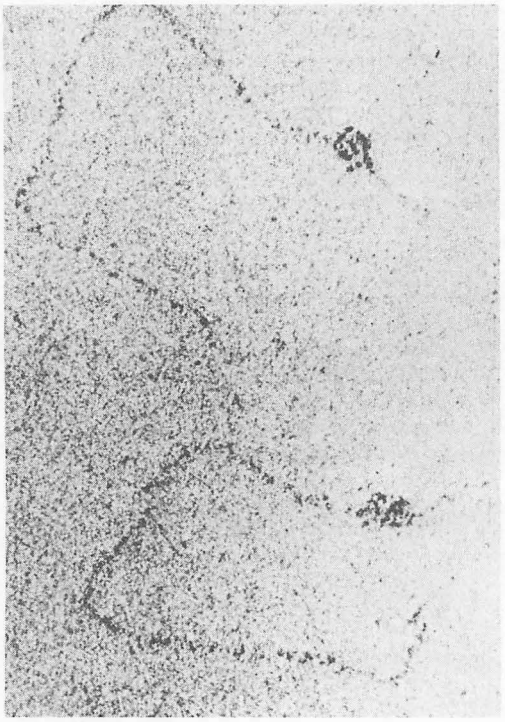


Figure 20.4
Electron microscope picture of an RNA polymerase molecule bound to a promoter.
RNA polymerase is bound to the pr promoter of bacteriophage **λ**. *The polymerase is located 110 base pairs from the end of a 970-base-pair DNA fragment.*
Reproduced with permission from J. Hirsch and R. Schlief, *Proc. Natl. Sci. USA*, 73:1518, 1976.

promoter regions, to which the RNA polymerase attaches. Through use of various techniques, including cloning and genetic engineering, discussed at the end of this chapter, the DNA from many operons can be isolated and the nucleotide sequence determined. The sites of attachment of the RNA polymerase have been determined by subjecting a mixture of RNA polymerase and the DNA of the operon to hydrolysis by DNase, and isolating the fragments of DNA that were not available to the hydrolytic action of the DNase. These fragments presumably were attached to and protected by the molecules of RNA polymerase (Figure 20.4). The nucleotide sequence of these DNA fragments can be determined, and the sequences of these presumed promoter regions (and surrounding areas) can be compared among various operons.

Figure 20.5*a* shows the nucleotide sequences of promoter regions of some bacterial operons. One would expect that promoter regions that recognize the same RNA polymerase would be identical or at least have base sequences in common. In addition, because the strength of interaction between the promoter region and the RNA polymerase varies from operon to operon, we might expect to see a pattern of nucleotide sequence that reflects these varying affinities. A computer analysis of these nucleotide sequences of promoter regions fulfills these expectations in a tentative way. In Figure 20.5*a*, the nucleotide sequences are arranged so that an apparently invariant T is lined up in each sequence. This T is the sixth nucleotide of heptamer TATAATG, called the *Pribnow box;* variations of this heptamer have been found in all bacterial and phage promoters. About 10 bases *downstream* (i.e., in the same direction that RNA polymerase moves during transcription; to the right in Figure 20.5*a*) is a circled nucleotide, which represents the initiation site for mRNA synthesis. This is determined by analyzing the sequence of RNA synthesized in vivo and in vitro.

Another region of conserved homology occurs about 15 bases *upstream* from the Pribnow box (i.e., in the opposite direction that the RNA polymerase moves during transcription; to the left in Figure 20.5*a*), or about 35 bases upstream from the initiation site. Allowing for some flexibility and other adjustments, there appears to be a significant homology between these promoters (Figure 20.5*b*). As shown, a similar region exists in the mammalian virus SV40, indicating that some mammalian RNA polymerases may have properties similar to the bacterial enzyme. There is some evidence that promoter regions with weak affinity for RNA polymerase have poor homology in this region. This area is thought to be the initial recognition site of the RNA polymerase prior to the formation of a more tightly bound complex of the RNA polymerase to the Pribnow heptamer.

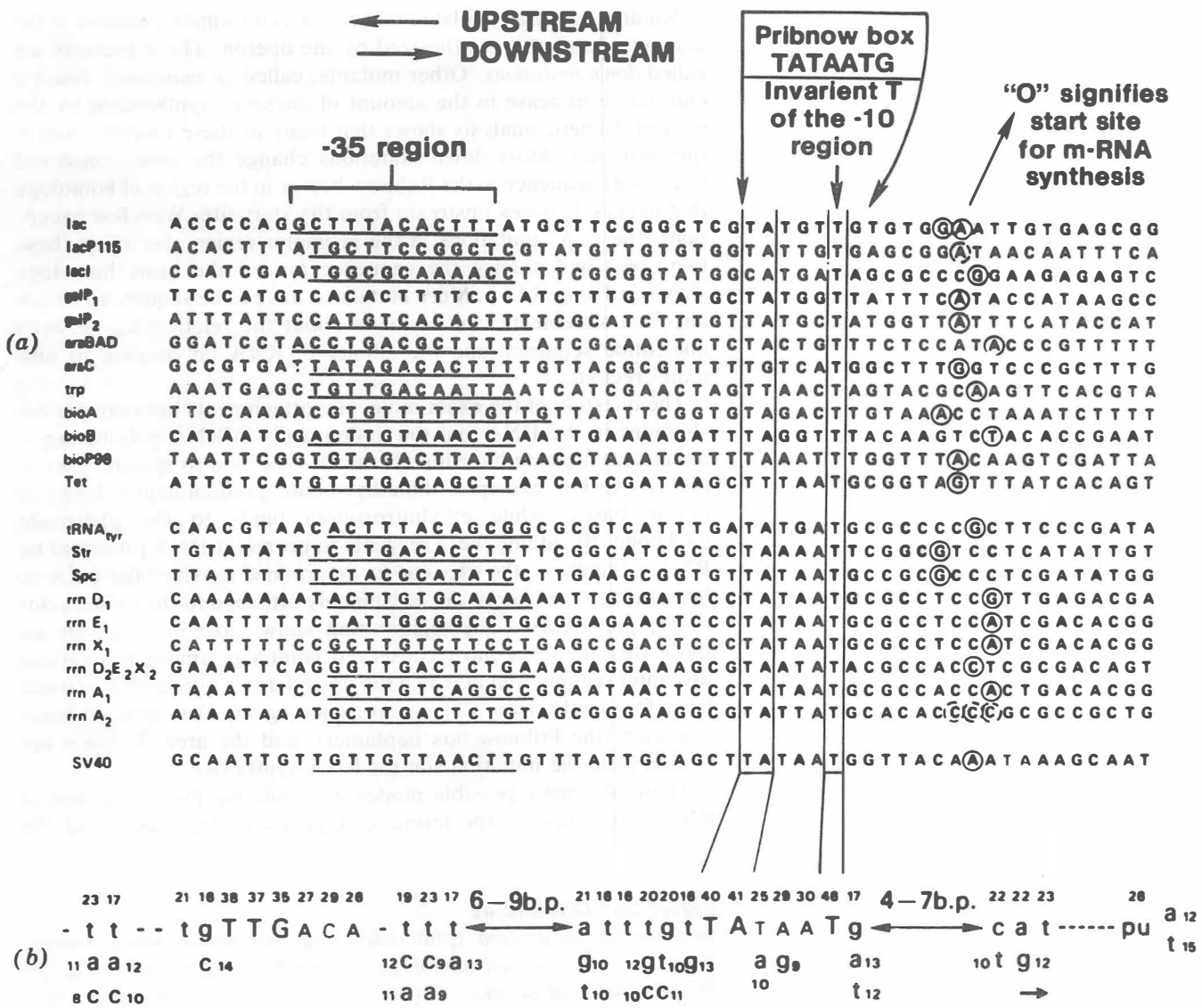


Figure 20.5
Sequences of some bacterial promoters and a theoretical prototypic sequence.
(*a*) *Promoter sequences. The antisense strand of the DNA is shown for each promoter. These are aligned with respect to an invariant T residue that occurs in the −10 region (i.e., the sixth position of the Pribnow sequence). This also aligns two other highly conserved −10 region positions. The circled base indicates the mRNA start site(s) for each promoter. The proposed −35 region sequence homologies are underlined.* (*b*) *Prototypic sequence. A statistical*

(Legend continued on p. 960)

Mutants have been isolated that cause a coordinate decrease in the amount of enzymes synthesized by the operon. These mutants are called *down mutations.* Other mutants, called *up mutations,* cause a coordinate increase in the amount of enzymes synthesized by the operon. Genetic analysis shows that many of these mutants map in the promoter. Most down mutations change the most conserved nucleotide sequence in the Pribnow box or in the region of homology that occurs 35 bases upstream from the start site. With few exceptions, the "up" mutations in the promoter region also affect these two areas and change the sequence *toward* the basic homology shown in Figure 20.5*b*. More mutations have to be sequenced before any firm statements can be made about the relationship between nucleotide sequence and the affinity of RNA polymerase to promoter regions.

The question of the exact molecular attachments between the nucleotides in the DNA and the amino acids in RNA polymerase is under investigation by using chemicals that bind to specific sites of the DNA. For example, dimethylsulfate predominantly binds to purine bases, while ethylnitrosourea binds to the phosphate backbone. By adding these reagents to promoter DNA protected by RNA polymerase, the sites can be determined to which the DNA no longer reacts and hence are presumably attached to the amino acids of the polymerase. The regions that show most involvement are those that show the most conserved homology among the various promoter regions, and in which the up and down promoter mutations most frequently occur. These are, once again, the area 10 bases upstream (the Pribnow box heptamer) and the area 35 bases upstream from the initiation site for RNA synthesis.

There are many possible modes of regulating the attachment of RNA polymerase to the promoter region and thus controlling the

***Figure 20.5* (*Continued*)**
summary of these data (plus other data not shown here) showing the extent of conserved homology of each base within the promoter. Highly conserved positions occurring in more than 75% of the cases are indicated by large uppercase letters; other well-conserved positions occurring in about 50% of the cases are indicated by smaller uppercase letters; positions conserved in less than 50% of the cases are indicated by lowercase letters. The number of promoter sequences having a particular conserved residue at each position is indicated. The mRNA start site for the prototypic promoter sequence is indicated.

subsequent rate and amount of transcription. For example, the rate of transcription might be fixed simply by the nucleotide sequence of the promoter region so that some operons have promoters with high affinity for RNA polymerase, leading to high rates of transcription, while some promoters have a low affinity for RNA polymerase leading to low rates of transcription. Alternatively, the availability of the promoter region to RNA polymerase might be modified by the attachment of regulatory proteins to the DNA, or regulatory molecules might attach to and modify the affinity of RNA polymerase. These types of regulation as well as others do exist and are described below.

20.4 THE CRP SITE: POSITIVE CONTROL OF TRANSCRIPTION INVOLVING CRP, A REGULATORY PROTEIN

The CRP site is present in most, if not all, operons that code for the enzymes involved in the utilization of various substances as carbon-energy sources, including the *lac* operon, the *ara* operon, the *gal* operon, and others. *E. coli* preferentially metabolizes glucose, and operons with CRP sites are turned off, or *repressed,* if glucose is available to the *E. coli* cell as a carbon-energy source. An example of this phenomenon is shown in Figure 20.6, where the appearance of β-galactosidase, one of the enzymes coded for by the *lac* operon, is delayed until all the glucose in the medium is depleted. Only then can lactose be used as the carbon-energy source. This phenomenon is called *catabolite repression,* because it occurs during the catabolism of glucose and may be due to a catabolite of glucose, rather than glucose.

Catabolite repression involves the effect of glucose on cAMP levels and a regulatory protein called *CRP* (*c*AMP *r*eceptor *p*rotein; also referred to as *CAP, c*atabolite *a*ctivator *p*rotein, or *CGA, c*atabolite *g*ene *a*ctivator). CRP is an allosteric protein, and when combined with cAMP is capable of binding to the CRP regulatory region of the *lac, ara, gal* and other operons near or adjacent to the site of binding of RNA polymerase (Figure 20.7). The combination of CRP and cAMP exerts positive control on transcription of these operons, since its attachment to the CRP site is necessary for the attachment of RNA polymerase (Figure 20.7*a*); alternatively, if the CRP site is unoccupied, the RNA polymerase cannot bind to the promoter site

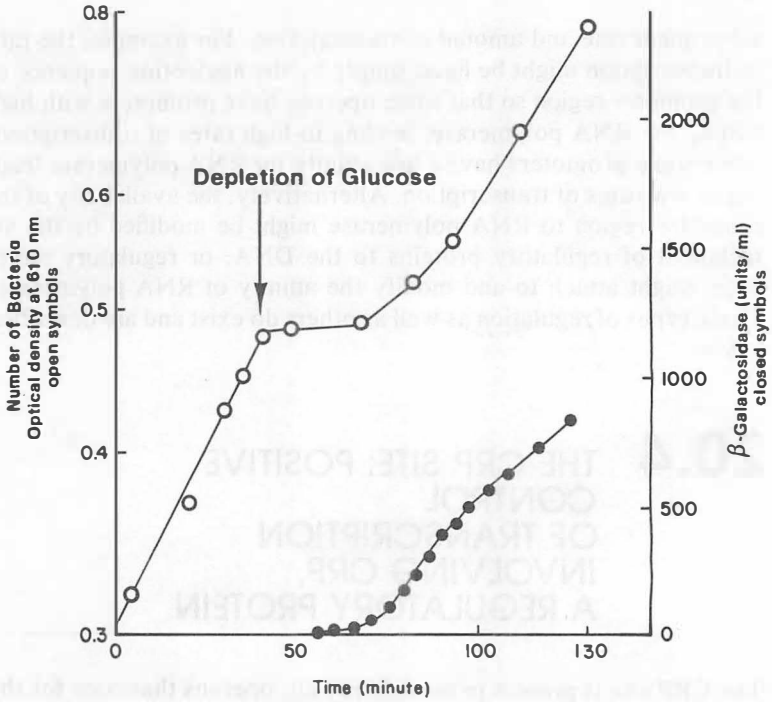


Figure 20.6
Lack of synthesis of β-galactosidase in* E. coli *when glucose is present.
The bacteria are growing in a medium containing initially 0.4 mg/ml glucose and 2 mg/ml lactose. The left-hand ordinate indicates the cell density of the growing culture, an indicator of the number of bacterial cells. The right-hand ordinate indicates the units of β-galactosidase per milliliter. Note that the appearance of β-galactosidase is delayed until the glucose is depleted.
Redrawn from W. Epstein, S. Naono, and F. Gros, *Biochem. Biophys. Res. Commun.*, 24:588, 1966.

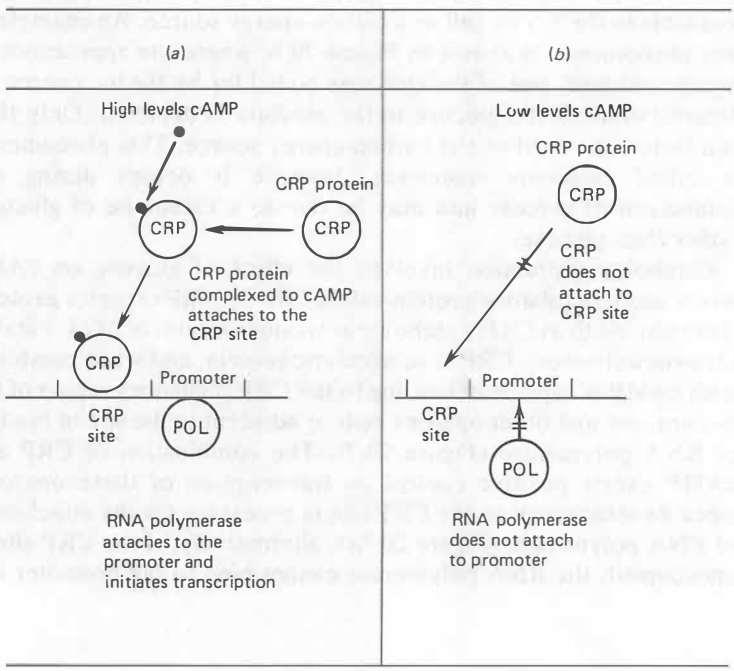


Figure 20.7
Schematic diagram showing how the level of glucose affects the transcription of the lac, gal, araBAD, and other operons.
(a) Low glucose level. (b) High glucose level. CRP = CRP protein; POL = RNA polymerase.

(Figure 20.7*b*). Glucose catabolism exerts its control because cAMP levels are low when glucose is present. The mechanism underlying this effect of glucose catabolism on cAMP levels is not known.

Thus transcription is controlled by glucose catabolism as follows: when the level of glucose catabolism is high, levels of cAMP are low, and cAMP is not available to combine with CRP (Figure 20.7*b*). The uncombined CRP does not attach to the CRP site, RNA polymerase does not attach to the promoter region, and transcription does not occur. The operon is repressed (turned off). On the other hand, when the levels of glucose catabolism are low, levels of cAMP are high, and cAMP is available to combine with CRP (Figure 20.7*a*). The CRP–cAMP complex attaches to the CRP site, RNA polymerase attaches to the promoter region, and transcription takes place. The operon is derepressed (turned on). Thus the level of glucose can regulate transcription of many operons through an interaction of cAMP with the regulatory protein, CRP, and the regulatory CRP site on the DNA of the *lac*, *araBAD*, *gal*, and other operons.

There is a great deal of evidence in favor of this mechanism of regulating the attachment of RNA polymerase to the promoter region. For example, cAMP levels are low in *E. coli* grown in glucose, and rapidly increase if glucose is not available. Addition of cAMP to *E. coli* growing in the presence of glucose overcomes catabolite repression of these operons. There is a good correlation between the level of cAMP and the amount of β-galactosidase synthesized in *E. coli*. Two classes of mutants exist that are not able to utilize many carbohydrates (e.g., lactose, arabinose, galactose) as carbon-energy sources. One group (Cya^-) has a defective adenylate cyclase and cannot synthesize cAMP from ATP. Addition of cAMP to the medium restores the ability of these bacteria to utilize the carbohydrates. The other group of mutants is not responsive to cAMP in the medium, and is lacking CRP (Crp^-). Genetically, the Crp^- mutants map in a different location of the genome than the Cya^- mutants.

Further proof that CRP and cAMP are involved is that these chemicals are essential for the initiation of *lac* mRNA synthesis in an in vitro (cell-free) system containing RNA polymerase, *lac* DNA, the four ribonucleoside triphosphates, and appropriate buffer. The use of an in vitro synthesizing system makes simpler the task of defining the components of a metabolic event, and it is used repeatedly to test and extend observations made in vivo. Observations of in vitro systems using DNA from *lac*, *gal*, *ara*, and other operons confirm that CRP and cAMP have a direct positive action on the transcription initiation of these many operons. It is not known whether CRP, when attached to the CRP site in the operon, has its major effect on the promoter DNA helix or on the RNA polymerase, or both.

There seem to be three types of promoters: those that have only a Pribnow box; those that have a Pribnow box and additional interaction sites around the region 35 bases upstream from the initiation site; and those that have the CRP site to help form a stable initiation complex between the RNA polymerase and promoter DNA. Of interest here are a series of mutants in the *lac* promoter which normally has a functional CRP site. A mutant has been found that has a defective CRP site and cannot initiate transcription of the *lac* operon. A "back"-mutation has been found that allowed initiation of *lac* transcription again. The nucleotide change, however, is not in the CRP site which is still defective; instead, two base pairs in the Pribnow box are changed so transcription could be initiated without CRP.

An interesting variation exists in the regulation of transcription of the *gal* operon. The *gal* operon has two overlapping promoters, each of which has a separate startsite for the initiation of transcription. Figure 20.8 shows the overlapping sequences of the Pribnow boxes

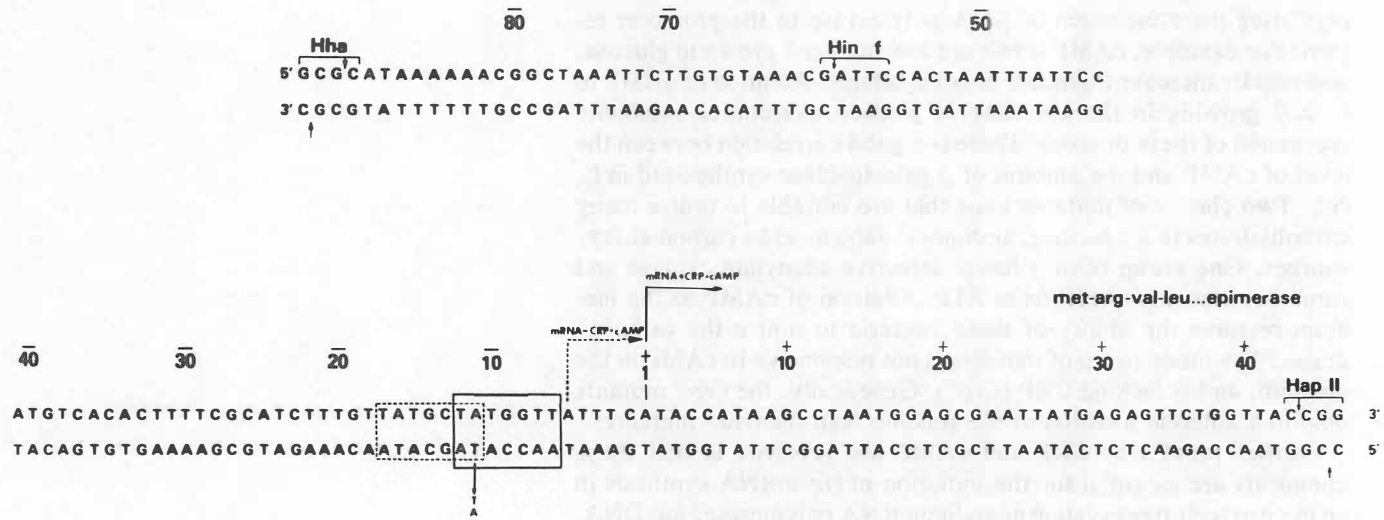


Figure 20.8

DNA sequence of the operator–promoter region of the galactose operon. *The ↑ (+1) corresponds to the start site of the promoter dependent on the CRP–AMP complex. The two Pribnow heptamers preceding each start site are boxed.*

Reproduced with permission from B. deCrombrugghe and I. Pastan, Cyclic AMP, the cyclic AMP receptor protein, and their dual control of the galactose operon, in *The Operon,* J. H. Miller and W. S. Reznikoff, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1978, p. 303.

of these two promoters and their separate start sites. One of these promoters is turned on by the CRP–cAMP complex, the other promoter is turned off by CRP–cAMP. The physiological importance of these two promoters depends on the fact that epimerase, an enzyme coded for by the *gal* operon, is necessary to convert glucose to UDP-galactose, a direct precursor for the synthesis of bacterial cell walls; hence this enzyme must be synthesized by the cell all the time. Thus the *E. coli* cell needs to maintain a low basal rate of synthesis for the epimerase coded for in the *gal* operon. This is accomplished by the CRP-independent promoter even in the presence of glucose.

The nucleotide sequence of the CRP site has been analyzed in DNA fragments protected by CRP from DNase (Figure 20.9). CRP sites vary in location and sequence. In the *lac* operon the CRP–cAMP complex attaches to a symmetrical region located some 55 to 70 bases upstream from the transcription initiation site. The *gal* op-

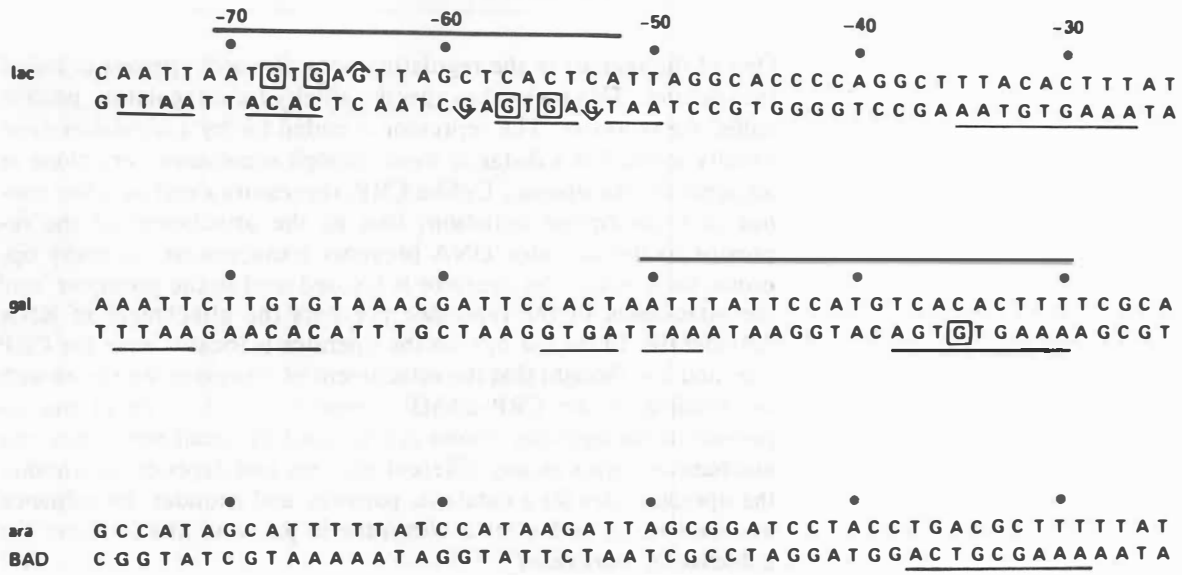


Figure 20.9
Nucleotide sequences in the presumed CRP sites of the lac, gal, and araBAD operons.
The overscored lines indicate regions thought to be recognized by CRP. Sequence homologies in lac, gal, and ara operons are underlined.

eron contains a region 50 to 70 bases upstream from the mRNA start site that is strikingly similar to the sequence in the *lac* operon, but the data indicate that a region between 30 and 50 bases upstream is also important to CRP interaction with the *gal* operon. Finally, the CRP site in the *ara* operon shows little homology to the CRP site in the *lac* or *gal* operons. Clearly the correlation of nucleotide sequence to the function of the CRP site is just beginning.

20.5 THE OPERATOR REGION: NEGATIVE CONTROL OF TRANSCRIPTION INVOLVING THE REPRESSOR, A REGULATORY PROTEIN

One of the regions in the regulatory area of many operons is called the *operator*. This region has specific affinity for a regulatory protein called the *repressor*. The repressor is coded for by a *regulatory gene* usually located at a distance from, though sometimes very close or adjacent to, the operon. Unlike CRP, repressors exert *negative control* of transcription initiation; that is, the attachment of the repressor to the operator DNA prevents transcription. In many operons, such as *lac*, the operator is located next to the promoter, and the attachment of the repressor prevents the attachment of RNA polymerase. In the *gal* operon the operator is located near the CRP site, and it is thought that the attachment of repressor interferes with the binding of the CRP–cAMP complex. The binding of the repressor to the operator region is regulated by small molecules; the mechanism varies among different operons and depends on whether the operon codes for a catabolic pathway and is under the influence of *induction,* or codes for a biosynthetic pathway and is under the influence of *repression*.

Induction

Operons that code for enzymes involved in the utilization or breakdown of compounds as carbon, energy, or nitrogen sources are usually *induced* or turned on by the substrate (or a close metabolite of the substrate) of the pathway. For example, the *lac* operon is induced (turned on) by lactose or allolactose, a metabolite of lactose, the *gal* operon is induced by galactose, and the *ara* operon by

arabinose, etc. Allolactose, the inducer of the *lac* operon, operates by combining with the repressor protein, which like CRP, is an allosteric protein, and its conformation and properties are changed when the inducer attaches. It is the uncombined repressor that attaches to the operator region and prevents the RNA polymerase from attaching to the promoter (Figure 20.10*a*). The repressor combined with inducer (allolactose) cannot attach to the operator region, and the RNA polymerase attaches to the promoter and initiates transcription (Figure 20.10*b*). A similar mechanism operates with respect to other operons.

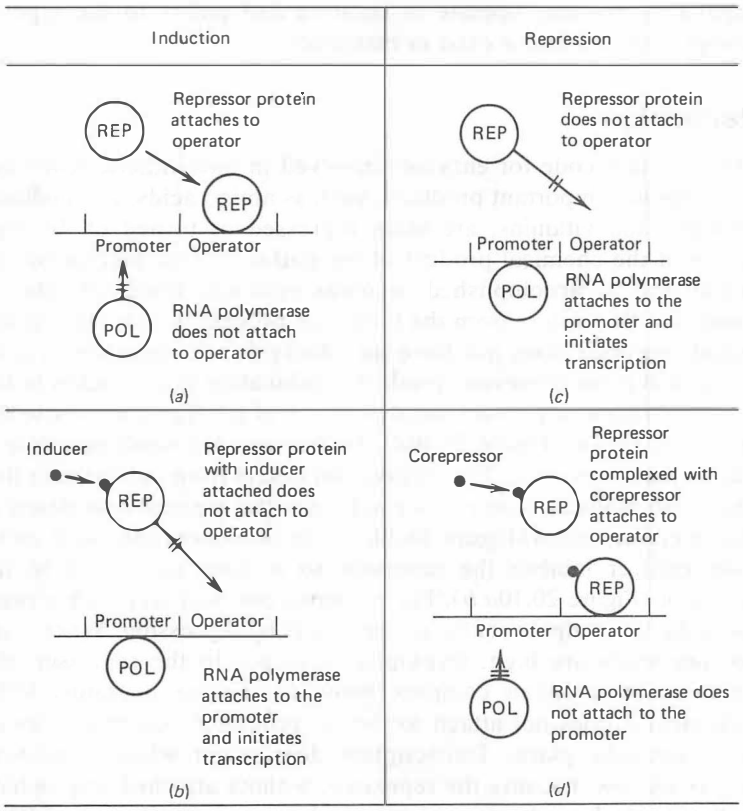


Figure 20.10
Schematic diagram showing control of transcription by operator–repressor interactions.
(a) Without inducer. (b) With inducer. (c) Without co-repressor. (d) With co-repressor. REP = repressor protein; POL = RNA polymerase.

An interesting variation in the function of the repressor exists within the *ara* operon, which is induced by arabinose. The repressor exerts negative control of transcription similar to the *lac* repressor, that is, arabinose combines with the *ara* repressor, and the complex does not attach to the *ara* operator. However, unlike the *lac* repressor, the *ara* repressor-arabinose complex also exerts positive control of transcription, and is required, in addition to the CRP–cAMP complex, for the initiation of transcription. Thus, the same regulator protein, the *ara* repressor, exerts negative control of transcription in the absence of arabinose when it combines with the *ara* operator region and positive control of transcription in association with arabinose. This example indicates the complex way some regulatory proteins operate in bacteria and points to the type of complexity that might exist in mammals.

Repression

Operons that code for enzymes involved in biosynthetic pathways that produce important products, such as amino acids, pyrimidines, purines, and vitamins, are often repressed or turned off by high levels of the chemical product of the pathway. One mechanism by which this is accomplished involves operator–repressor interactions, but this differs from the induction process in that the uncombined repressor does not have an affinity for the operator region; instead, it is the repressor–product combination that attaches to the operator region to prevent the attachment of RNA polymerase to the promoter region (Figure 20.10*d*). In this case the small molecule is called the *corepressor*. Thus repression differs from induction in that the small molecule corepressor activates the repressor to attach to the operator region (Figure 20.10*c*,*d*); in induction, the small molecule inducer inhibits the repressor so it does not attach to the operator (Figure 20.10*a*,*b*). For example, one way tryptophan regulates the transcription of the *trp* operon is by repression. When tryptophan levels are high, tryptophan attaches to the repressor, the repressor–tryptophan complex binds to the *trp* operator, RNA polymerase does not attach to the *trp* promoter, and transcription does not take place. Transcription does occur when tryptophan levels are low, because the repressor, without attached tryptophan, will not attach to the *trp* operator and prevent attachment of RNA polymerase.

Repressors Affecting More Than One Operon

CRP is a regulatory protein affecting many operons, whereas repressors are regulatory proteins that usually affect only one operon.

There are instances, however, where repressors have been reported that exert negative control of transciption on more than one operon. These examples are significant in that similar phenomena may occur in mammals. For example, it has been reported that the enzymes for the breakdown and utilization of histidine in the bacterium *Klebsiella aerogenes* are coded for by two operons, each with a separate operator region that is affected by the same repressor. There is evidence that the *trp* repressor also acts on *arom H*. Similarly, there are two operators in phage λ that are controlled by the same repressor synthesized by the regulator gene C_1. This case is interesting because the nucleotide sequence of each operator is different. Thus this repressor appears to be able to recognize and bind to different nucleotide sequences in the DNA. It is not known whether similar phenomena exist in mammals.

Mutants Affecting Repressor–Operator Interactions

There are large numbers of mutants that affect the interactions between the repressor and the inducer or corepressor, and those between the repressor and the operator region. Mutants affecting the *lac* operon will be described here. In general, similar mutants exist with respect to other operons where similar operator–repressor mechanisms exist. Where variations from the induction system of the *lac* operon exists, such as the positive control by the repressor of the *ara* operon or repression of the *trp* operon, corresponding differences in the types of mutants are found. Mutant analysis is very valuable because in many instances the analysis of the biochemical and genetic properties of mutants, and their interactions with other mutants, provided the first clues to the mechanisms of the underlying reactions and led to the first correct theoretical explanations of the control of transcription and other phenomena by Jacob and Monod, and others. In addition, mutants are proving useful in correlating molecular structure (nucleotide sequence) and function, since they provide altered nucleotide sequence for analytical comparison.

Mutations of the *lac* operon structural genes, *Z, Y,* or *A*, affect only the enzyme coded for by that structural gene. Mutations that lead to the absence of enzyme activity are most easily detected, since such mutants are not able to grow on lactose as a sole carbon-energy source. Often the inactive enzyme in such mutants can be detected as an antigen, called *CRM* (cross reacting material). In contrast, mutants located in the regulatory regions of the *lac* operon (promoter or operator regions) affect the synthesis of all three enzymes of the *lac* operon *coordinately;* indeed it is this *coordinate control* of enzyme synthesis that supports the idea that the operon constitutes the basic

unit of transcription. As discussed above, mutations in the *lac* promoter region decrease or enhance the synthesis of the three proteins by decreasing or enhancing the attachment of RNA polymerase either by directly changing key nucleotide sequences, or indirectly, by changing the CRP site.

The mutations found at the *lac* operator region are of two types: operator-constitutive (O^c) or operator-negative (O^0). Strains carrying the O^c mutation produce moderate levels of the three enzymes coded for by the *lac* operon, even in the absence of lactose. This phenomenon of synthesizing enzyme in the absence of inducer is called *constitutive synthesis.* It is as if the *lac* O^c mutation produced a decreased affinity between the operator region and the repressor, and the repressor was not attaching as frequently to the operator region to prevent the RNA polymerase from attaching to the promoter region to initiate transcription. Since full induction still requires the presence of lactose, there must still exist some affinity of the operator region for the repressor in these mutants.

Strains carrying the O^0 mutation do not synthesize the three enzymes coded for by the *lac* operon, even in the presence of lactose. This suggests that the nucleotide sequence in the *lac* operator region is altered so the affinity for the repressor is increased, and the repressor attaches even if it is combined with lactose. Thus changes in the nucleotide sequence of the *lac* operator region alter the affinity of the operator for the repressor; the O^c mutants lower the affinity, while the O^0 mutants raise the affinity. This in turn affects the binding of the RNA polymerase to the *lac* promoter region and affects transcription.

Mutations of the regulatory gene *i* that codes for *lac* repressor affect the affinity of the repressor protein for the operator or for lactose, and thereby affect the transcription of the *lac* operon and ultimately the ability of the cell to grow on lactose. Mutations of the *i* gene are mainly of two types; the absence of functional repressor (i^-) and the presence of superrepressor (i^s), that is, a repressor not neutralized by lactose. Strains containing the i^- mutants produce high levels of the three enzymes coded for by the *lac* operon *constitutively,* that is, even in the absence of lactose. This suggests that these strains contain no repressor or a repressor with an altered operator binding site, so the molecule does not bind to the *lac* operator region, and RNA polymerase is not prevented from binding to the *lac* promoter region and initiating transcription. Strains containing the i^s gene contain little or none of the three proteins coded for by the *lac* operon; it seems that the structure of the lactose binding site of the repressor protein is altered and lactose cannot bind and thus prevent attachment of the repressor to the *lac* operator region. This "super" repressor is thus free to always bind to the *lac* operator region, and

transcription is permanently turned off. There are also i^r mutants, in which there is increased binding between *lac* repressor and *lac* operator DNA, and decreased binding between *lac* repressor and inducer.

There are considerable biochemical data that indicate the correctness of these interpretations of these mutants. *Lac* operator regions can be isolated by protecting the DNA from DNase with *lac* repressor protein. Such operator regions have the expected properties. For example, *lac* operator DNA isolated from normal strains has a high affinity for the *lac* repressor; such affinity is greatly reduced if lactose is present. If the *lac* repressor is isolated from an i^s mutant, lactose has little effect on protein-DNA binding, as expected. If the *lac* operator DNA is isolated from an O^c mutant strain, the affinity for *lac* repressor is low, whereas if the operator DNA is from an O^0 mutant, the affinity of the repressor for the DNA is high, and relatively unaffected by lactose. This agreement between the affinity in vitro between various *lac* repressor proteins and *lac* operator DNA from wild-type and mutant strains, and the effects of lactose, is consistent with the interpretations of the phenomena observed in mutant strains.

Correlating Molecular Structure and Function of Operator Regions

Figure 20.11*a* shows the nucleotide sequence of the *lac* operon DNA protected from enzymatic hydrolysis by *lac* repressor; presumably this DNA represents the *lac* operator. This fragment of DNA contains about 27 to 28 base pairs and is characterized by twofold symmetry. This is of interest in view of the structure of the *lac* repressor protein which has an elongated shape, whose cross section looks like a dumbbell (Figure 20.12), suggesting that the *lac* repressor binds to the lac operator DNA so that the symmetry of operator region corresponds to the symmetry of the repressor. However, the data on the effect of *lac* repressor on chemical modification of the DNA of the *lac* operator region, summarized in Figure 20.11*b*, suggest otherwise. These experiments determine the effect of *lac* repressor binding to the *lac* operator DNA on the reactivity of the chemical constituents of the DNA to various chemical treatments. Figure 20.11*b* shows the sites that are affected by the *lac* repressor binding to *lac* operator. Thus the binding of *lac* repressor inhibits the methylation of seven purines, while enhancing the methylation of three purines by dimethyl sulfate; inhibits thymine cross-linking at 13 sites in BrU-substituted *lac* operator DNA; and inhibits alkylation of seven phosphate groups by ethylnitrosourea. A detailed analysis of these data suggests that the *lac* repressor binds along one side of the

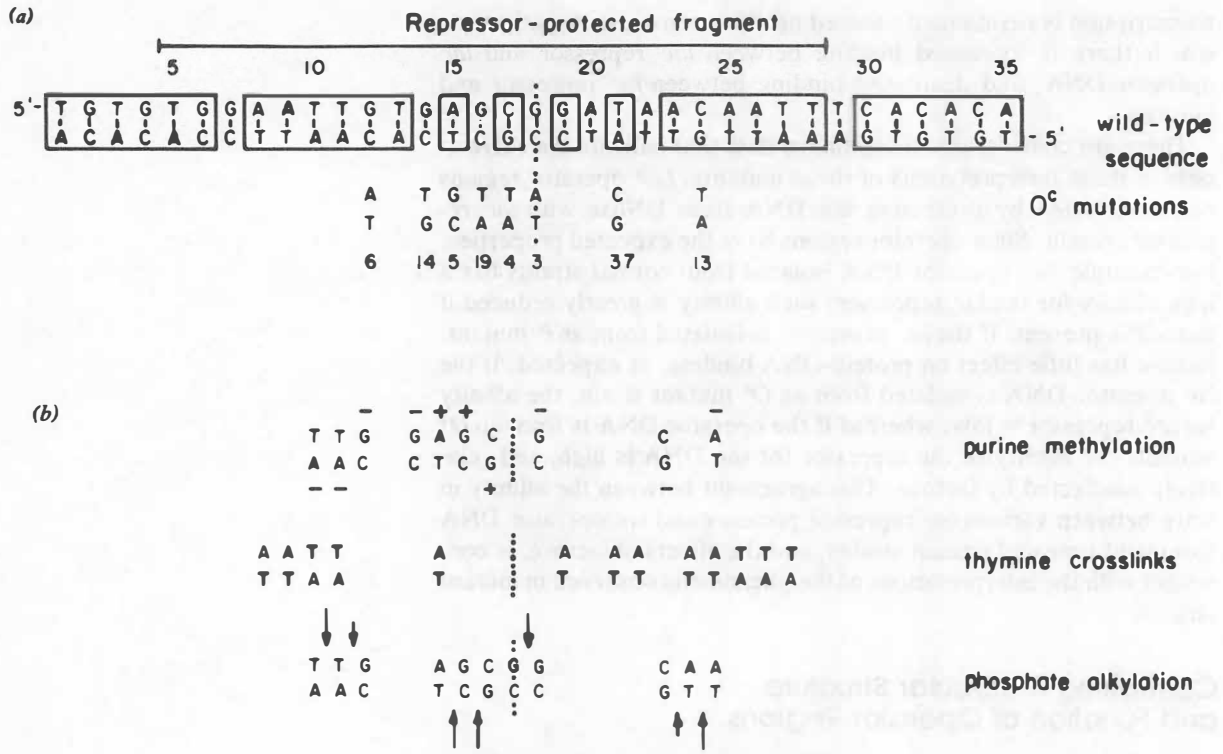


Figure 20.11

Nucleotide sequence of the lac operator region showing (a) nucleotide changes produced by O^c mutations, and (b) nucleotides protected by lac repressor from various chemical treatments.

(a) Sequence of the lac operator. The 35-base-pair sequence of the wild-type operator region; 28 base pairs (enclosed in boxes) are symmetric about a twofold axis (vertical dotted line). The DNA fragment protected from nuclease digestion by lac repressor is indicated by the solid line. The lower portion shows the base-pair substitutions of O^c mutations. (b) Chemical modifications of lac operator DNA. The upper portion shows the positions at which lac repressor inhibits (−) or enhances (+) methylation of purines by dimethyl sulfate. The middle portion shows the positions of the thymines at which repressor cross-links to BrU-substituted DNA upon uv irradiation. The lower portion shows the positions at which repressor inhibits alkylation of phosphates by ethylnitrosourea, to a greater (long arrows) or lesser (short arrows) degree.

Reproduced with permission from M. D. Barkley and S. Bourgeois, (1978) Repressor recognition of operator and effectors, in *The Operon*, J. H. Miller and W. S. Reznikoff, eds., Cold Spring Harbor Laboratory, N.Y., 1978, p. 177.

helix, makes contact in both the major and minor grooves of the *lac* operator DNA, and interacts more strongly with the left half of the operator. Other studies involving substitutions of various nucleotides in the *lac* operator are consistent with this tentative conclusion, and reveal asymmetry in the interaction of the *lac* repressor and *lac* operator DNA. Thus it is not known how much of the symmetry observed in the *lac* operator is important to its interaction with the repressor.

20.6 REGULATION OF TRANSCRIPTION BY ALTERING THE STRUCTURE OF RNA POLYMERASE

The regulation of transcription by changing the structure of the RNA polymerase is not common in bacteria, but it is found among phage that grow in bacteria and may exist in some viruses that grow in mammalian cells. Many bacterial phages exhibit a well-defined temporal sequence for transcription of its genes. For example, phage *SP01* uses the bacterium *Bacillus subtilis* as its host. Following infection, the bacterial host RNA polymerase attaches to the promoters of phage genes that are transcribed during early phage development; the bacterial host RNA polymerase does not recognize the promoters for genes that are transcribed later. One of the phage genes transcribed early in development is a regulatory gene that produces a protein that binds to the bacterial host RNA polymerase and changes its properties. As a result, the altered host RNA polymerase can attach to the promoters of the phage genes for the middle stages of phage development. The protein products of two other phage regulator genes are synthesized during the middle stages of development. These attach to the bacterial host RNA polymerase, which is able to initiate transcription of the phage genes required for late development of the phage.

We can hypothesize that the nucleotide sequences of the promoter differ between early, middle, and late developmental phage genes. In order for middle and late developmental phage genes to be transcribed, the structure of the bacterial RNA polymerase must be changed for the host enzyme to recognize phage gene promoters. This change is accomplished by proteins coded for by the phage in appropriate developmental sequence; in this way a temporal sequence is built into phage development. Figure 20.13 shows the

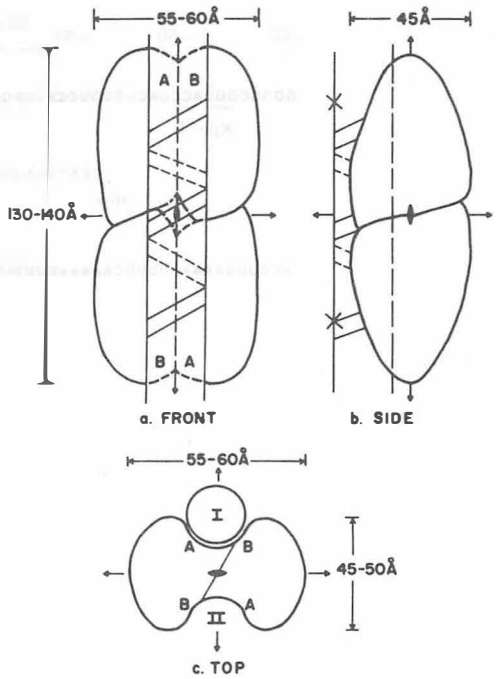


Figure 20.12
Diagrammatic representation of the lac repressor tetramer indicating a possible mode of interaction with the operator.
Each of the four subunits has two different surfaces, A and B, which can interact with the DNA, resulting in the possibility of two operator binding sites (I and II) per repressor.
Reproduced with permission from T. A. Steitz, T. J. Richmond, D. Wise, and D. Engelman, *Proc. Natl. Acad. Sci. USA*, 71:593, 1974.

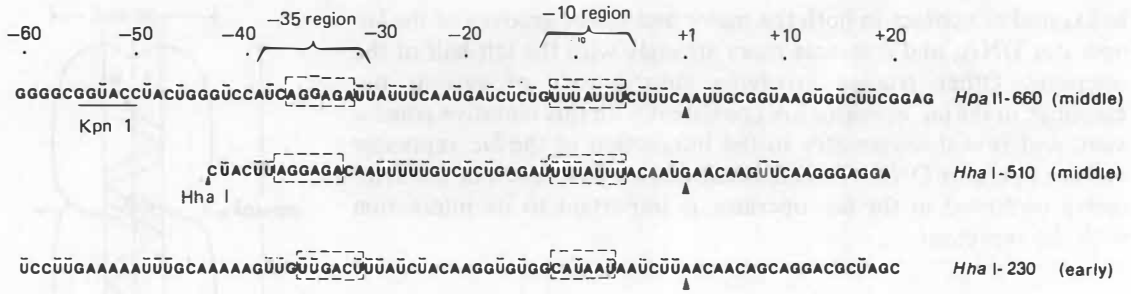


Figure 20.13
Nucleotide sequences of two middle gene promoters and one early gene promoter of phage SP01.
The sequences were aligned according to the initiation sites for RNA polymerase denoted by large vertical arrowheads. The identical sites in the −10 and −35 regions are enclosed in boxes.
Reproduced with permission from C. Talkington and J. Pero, *Proc. Natl. Acad. Sci. USA*, 76:5465, 1979.

nucleotide sequences of the promoters of two middle developmental phage genes and one early developmental phage gene. The two middle genes are identical within the boxed regions, and both are clearly different from the early gene promoter. More data are needed to compare the nucleotide sequences of late developmental genes, as well as additional early and middle developmental genes, before firm conclusions can be made. It does seem likely, however, that the phage achieves the appropriate temporal pattern in the transcription of its early, middle, and late developmental genes by different promoter sequences and by coding for regulatory proteins that change the properties of host RNA polymerase. It is not known whether similar processes occur in mammalian cells.

20.7 THE ATTENUATOR REGION: REGULATION OF THE TERMINATION OF TRANSCRIPTION

The regulatory role of the *attenuator region* seems to be the control of termination of transcription prior to the transcription of the structural genes. This regulatory region has been reported in operons

involved in the synthesis of the amino acids, tryptophan, leucine, histidine, phenylalanine, and threonine. Operons coding for the enzymes synthesizing other amino acids may also contain this regulatory region. It is important to note that the *his* operon, and possibly the *leu, thr,* and *phe* operons, lack operator regions, and the attenuator may be the major control of transcription in these operons.

The attenuator is located between the initiation site for transcription and the first structural gene. Thus transcription gives rise to an RNA transcript, called *leader RNA*. Termination of transcription at the attenuator site is responsive to the level of the appropriate amino acid. For example, if the level of tryptophan is in excess of the needs of the cell, termination of transcription at the attenuator of the *trp* operon occurs. Transcription termination does not take place if the level of tryptophan is low.

The nucleotide sequence of the leader RNA or of the DNA in the region preceding termination provides important information on how termination of transcription might occur. Figure 20.14 presents the sequence for the *trp* operon; the *his, phe, thr,* and *leu* operons have similar sequences. Each has a sequence for ribosome binding and peptide chain initiation followed by codons for 14 to 28 amino acids. Among these codons are multiple tandem codons for the amino acid related to the operon. For example, out of 16 codons, the leader RNA of the *his* operon has 7 histidine codons in tandem; out of 14 codons, the *phe* operon leader RNA has 7 phenylalanine codons, though not all in tandem; out of 21 codons, the *thr* operon (which responds to the level of isoleucine *and* threonine) has 8 threonine codons and 4 isoleucine codons, 11 of which are in tandem in the leader RNA; out of 28 codons, the *leu* operon leader RNA has 4 leucine codons in tandem, and out of 14 codons, the *trp* operon leader RNA has 2 tryptophan codons in tandem. Following the codons, there exist signals for the termination of translation. Finally,

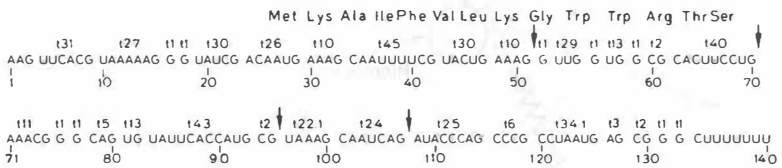


Figure 20.14
Nucleotide sequence of leader RNA from the trp operon.
The 14 amino acids of the putative leader peptide are indicated over their codons.
Reproduced with permission from D. L. Oxender, G. Zurawski, and C. Yanofsky, *Proc. Natl. Acad. Sci. USA,* 76:5524, 1979.

there exists, before the start of the first structural gene of the operon, a region with a nucleotide sequence similar to transcription terminators at the end of operons; that is, a region of dyad symmetry and a run of GC base pairs followed by a run of A's in the DNA coding strand so the RNA will contain a run of U's (see Chapter 18).

Based on this nucleotide sequence, analyses of mutant strains, in vivo and in vitro transcription systems, and other data, the following model has been generated to explain how the level of an amino acid can affect transcription termination in the attenuator region of these operons (Figure 20.15). Following attachment to the promoter region, RNA polymerase initiates transcription of the DNA segment that exists prior to the first structural gene. The transcription of this region produces what is called *leader RNA*. A ribosome attaches to the ribosome binding site of the leader RNA and translation of the codons into the leader peptide takes place before transcription of the leader RNA is finished. Whether translation proceeds smoothly depends on the levels of the appropriate amino acid and its charged tRNA. For example, if the level of tryptophan is high, then the level of charged tryptophan-tRNA is high, and the ribosome has no difficulty in translating the *trp* operon leader RNA. However, if the levels of tryptophan and charged tryptophan-tRNA

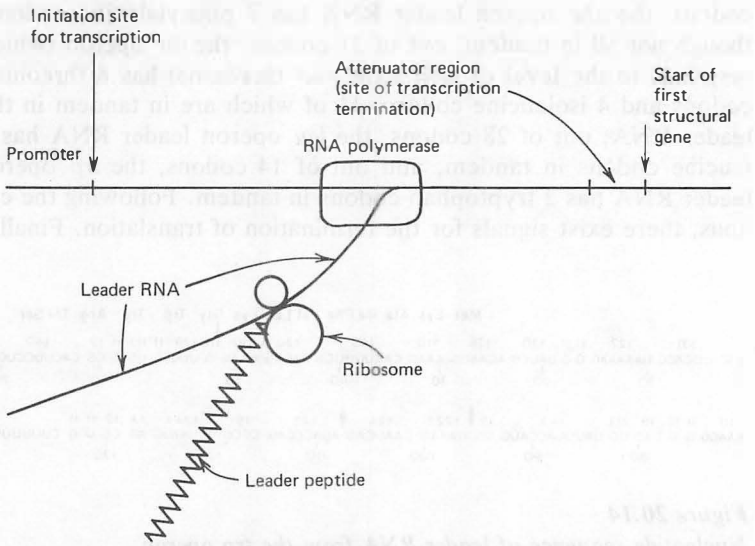


Figure 20.15
Schematic representation of the formation of the leader RNA and the leader peptide involved in attenuation.

is low, the ribosome will stall at the two contiguous *trp* codons on the leader RNA. This stalling affects the secondary structure of the leader RNA and provides the signal that prevents termination of transcription, and the RNA polymerase reads through the termination point into the first structural gene.

To understand how a stalled ribosome can prevent termination of transcription at a point more than 100 nucleotides downstream, it is necessary to examine the potential secondary structure of the leader RNA. This RNA contains considerable symmetry that allows the formation of several alternative stem and loop structures. Figure 20.16 depicts some stem and loop structures possible for the *trp* operon leader RNA; similar secondary structures exist in the leader RNA of the *leu, phe, thr,* and *his* operons. It can be seen in Figure 20.16 that four regions of base pairing exist and these can give rise to three stem and loop structures. Region 1 can pair with region 2, region 2 with 3, and region 3 with 4. The model proposes that pairing between regions 3 and 4 is a signal to terminate transcription. Figure 20.17 shows how excess levels of tryptophan, facilitating ribosomal movement across the two tryptophan codons, allows the formation of the 3-4 stem and loop, whereas low levels of tryptophan would, by causing the ribosome to stall, allow the formation of a 2-3 stem and loop. The latter prevents the formation of the 3-4 stem and loop, and termination does not take place. In the absence of any translation by the ribosome (as occurs in mutants that change the ribosome binding site), the 1-2 stem and loop would form, followed by the 3-4 stem and loop, resulting in termination.

There are considerable genetic and biochemical data in agreement with this model. As expected, mutants that affect the level of tRNA or its charging affect termination at the attenuator and thus the level of enzymes coded by the operon. Such mutants change the nucleotide sequence of the tRNA itself, the synthetase enzyme required to activate the amino acid prior to charging, or enzymes that modify the tRNA after it is formed. In addition, mutants that change the nucleotide sequence of the leader RNA so ribosome binding or the formation and stability of the stem loops is affected also change termination rates with subsequent changes in enzyme activities coded by the operon. Deletions of the attenuator region would be expected to totally eliminate termination and, in fact, deletions of *trp* operon attenuator DNA produce strains with levels of the enzymes coded for by the *trp* operon about 8- to 10-fold higher than usual.

The model is also consistent with information obtained using an in vitro system to synthesize *trp* operon leader RNA. This system contains RNA polymerase, a DNA template containing the *trp* operon, and the four nucleoside triphosphates. Termination in this system is very efficient, and a leader RNA about 140 nucleotides long is pro-

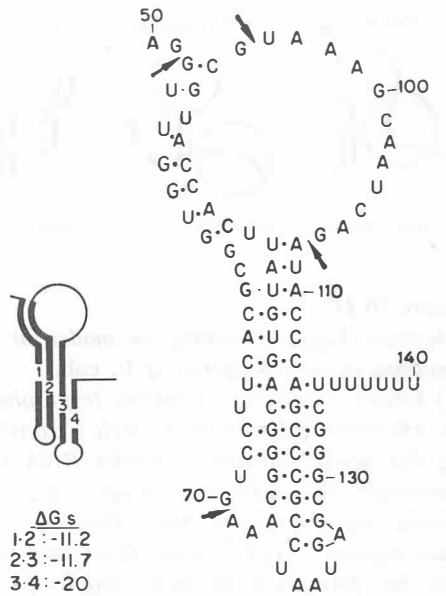


Figure 20.16
***Schematic diagram showing the proposed secondary structures in trp leader RNA from* E. coli.**
Four regions can base-pair to form three stem and loop structures. These are shown as 1-2, 2-3, and 3-4.
Reproduced with permission from D. L. Oxender, G. Zurawski, and C. Yanofsky, *Proc. Natl. Acad. Sci. USA*, 76:5524, 1979.

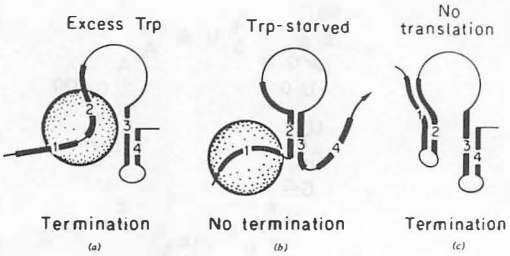


Figure 20.17
***Schematic diagram showing the model for attenuation in the trp operon of* E. coli.**
(a) Under conditions of excess tryptophan, the ribosome (the shaded circle) translating the newly transcribed leader RNA will synthesize the complete leader peptide. During this synthesis the ribosome will mask regions 1 and 2 of the RNA and prevent the formation of stem and loop 1-2 or 2-3. Stem and loop 3-4 will be free to form and signal the RNA polymerase molecule (not shown) to terminate transcription. (b) Under conditions of tryptophan starvation, charged tryptophan-tRNA will be limiting, and the ribosome will stall at the adjacent trp codons in the leader peptide coding region. Because only region 1 is masked, stem and loop 2-3 will form, excluding the formation of stem and loop 3-4, which is required as the signal for transcription termination. Therefore RNA polymerase will continue transcription into the structural genes. (c) Under conditions in which the leader peptide is not translated, stem and loop 1-2 will form preventing the formation of stem and loop 2-3, and thereby permit the formation of stem and loop 3-4. This will signal transcription termination.
Reproduced with permission from D. L. Oxender, G. Zurawski, and C. Yanofsky, *Proc. Natl. Acad. Sci. USA*, 76:5524, 1979.

duced. A similar leader RNA has been isolated from cells under derepressed conditions. Experiments on this leader RNA using RNase to detect RNase-resistant base pairing in stem and loop structures reveals extensive secondary structure similar to the ones predicted on the basis of this model (Figure 20.16).

Thus the mechanism of control at the attenuator region differs from the CRP and the operator sites where regulatory proteins attach to the DNA, depending on the presence of crucial small molecules. Regulation of transcription at the attenuator utilizes the nucleotide sequence and the secondary structure of the transcribed leader RNA; the ribosome mimics the role of regulatory protein and exerts *positive* control of transcription during the translation of the leader RNA; and the controlling small molecule is the specific tRNA charged with the amino acid that is the product of the enzymes coded by the operon.

N Protein from λ Phage Is an Antiterminator

The first step in the development of λ phage is the transcription of immediate-early developmental genes. During the early stages, transcription of phage genes lying beyond these genes is prevented or reduced by transcription termination. The N protein, a product of one of the immediate-early phage genes, acts as an antiterminator, and in its presence the RNA polymerase reads through these transcription termination signals of λ phage. The N protein also operates as an antitermination agent at a variety of sites in λ phage and some *E. coli* operons; unexpectedly, the promoter to which the RNA polymerase attaches is of great importance for N protein effectiveness. The mechanism of action of protein N is not known. It might act by interacting with RNA polymerase so it does not terminate. Alternatively it could bind to the nascent RNA and interfere with the stem and loop formation required for transcription termination or it might act to stall ribosomes as in the model of attenuator control.

20.8 STRINGENT CONTROL OF THE SYNTHESIS OF RIBOSOMAL AND TRANSFER RNA

When *E. coli* and other bacteria are starved for amino acids, widespread changes occur in RNA synthesis, particularly a decrease in

the rate of synthesis of rRNA and tRNA. This effect of starvation on the transcription of these genes is referred to as *stringent control.* The effects on the synthesis of mRNA and on enzyme protein is varied; many increase, many decrease, most remain unaffected. Stringent control is relaxed in a mutant, *relA*, in which two unusual guanine nucleotides are missing. The two nucleotides are guanosine 5′-diphosphate-3′-diphosphate (ppGpp) and guanosine 5′-triphosphate-3′-diphosphate (pppGpp). The *relA* gene product is an enzyme attached to ribosomes; the attachment of uncharged tRNA to the ribosome stimulates the synthesis of these two nucleotides and appears to be the signal for stringent control. Thus starvation leads to low levels of amino acids, which in turn leads to increased levels of uncharged tRNA. Both ppGpp and pppGpp appear capable of affecting the transcription of many operons, yet others are unaffected. The mechanism of how this occurs is unknown, but it might be analogous to the way cAMP interacts with CRP and affects the attachment of RNA polymerase to promoter regions. Thus an unidentified regulatory protein (analogous to CRP) that combines with ppGpp or pppGpp may exist. A careful comparison of the nucleotide sequences of the regulatory regions of operons that are and are not affected by stringent control might shed light on this phenomenon.

20.9 CONTROL OF TRANSCRIPTION OF VARIOUS OPERONS IN BACTERIA

Several ideas should now be clear. First, the similarities and differences in the regulation of transcription of different operons of bacteria clearly reflect the physiological and biochemical roles of the substrates and products of the metabolic pathways involved. Second, the regulation of transcription of operons (and thus of the synthesis of the enzymes coded for by the operon) is often ultimately carried out by small metabolites in the cell, inducers, corepressors, cAMP, tRNA, ppGpp, pppGpp, and so on. These chemical signals are effective on specific regulatory proteins that regulate transcription by affecting the attachment of the RNA polymerase to the promoter region, or by termination of transcription at the attenuator region. There are probably regulatory proteins and small molecules effectors yet to be recognized. The bacterial cell in this way transfers important chemical information about its internal mileu to the DNA

and to RNA polymerase. The regulating molecules may in one sense be viewed as intracellular hormones. What follows is a short summary of how transcription in some bacterial operons is regulated (see Figure 20.3).

The lac Operon

This operon codes for three enzymes involved in the transport and hydrolysis of lactose and thus in the utilization of lactose as a carbon/energy source when glucose is deficient. The regulation of transcription, and consequently the synthesis of the proteins coded for by the structural genes of the operon, is accomplished by the levels of lactose and glucose and by the nucleotide sequence of the promoter. Lactose is converted to allolactose; this metabolite induces the *lac* operon by interacting with the regulatory protein, the *lac* repressor. Glucose affects transcription of the *lac* operon via catabolite repression involving cAMP and CRP. The structure of the promoter region determines the ability of the RNA polymerase to attach to the promoter DNA and to initiate transcription at a fixed rate, providing all other regulatory signals are turned on. There is no attenuator region.

The araBAD Operon

The breakdown and utilization of arabinose as a carbon-energy source requires five structural genes, three of which, *araB, araA, araD,* are located in a single operon. Two other structural genes, *araE* and *araF*, are located separately elsewhere in the *E. coli* genome. The repressor, a regulatory protein coded for by *araC,* can attach to the operator for the *araBAD* operon, as well as the operators controlling *araE* and *araF*. Transcription of the *araBAD* operon is ultimately regulated by the levels of two small molecules, arabinose and glucose. Arabinose operates as an inducer and combines with the regulatory protein, the repressor. Such a combination has two effects; first, the repressor is prevented from attaching to the operator region, and second, the arabinose–repressor combination acts in a positive way to facilitate, along with CRP–cAMP, the attachment of the RNA polymerase to the promoter. The mechanism of this effect is not clear. Glucose controls transcription through catabolite repression involving cAMP and CRP. The nucleotide sequence of the promoter affects the affinity for RNA polymerase, and thus the rate of transcription. There is no attenuator region.

The gal Operon

The *gal* operon codes for three enzymes involved in the conversion of galactose to UDP-glucose. This pathway has two metabolic func-

tions: to utilize galatose as a carbon-energy source when galactose is present in the absence of glucose, and to produce UDP-glucose for cell wall synthesis for which there is continuing need. The need for a sporadic high level of enzyme synthesis and a constant low level of enzyme synthesis is resolved by the presence of two overlapping promoters, one of which operates when low levels of enzymes are needed to make UDP-glucose from glucose for cell wall synthesis, the other when galactose is available as a carbon/energy source in the absence of glucose. The level of glucose regulates transcription through catabolite repression involving cAMP and CRP. The cAMP–CRP complex has two effects; if it is not attached to the CRP site of the operon (glucose level is high), the promoter for galactose utilization is turned off, and the promoter for low-level transcription is turned on. If CRP–cAMP is attached to the CRP site (glucose level is low), the promoter for high-level transcription is turned on and the low-level promoter is turned off. The level of galactose regulates transcription as an inducer that combines with the *gal* repressor to prevent attachment of the repressor to the operator. Because the operator region overlaps the CRP site, it is thought that the *gal* repressor interferes with the attachment of CRP–cAMP rather than the RNA polymerase, although this is not known with certainty. The nucleotide sequence of the promoter affects the affinity for the RNA polymerase, and thus the rate of transcription. There is no attenuator region.

The trp Operon

The *trp* operon codes for four enzymes in the conversion of the chorismate to tryptophan. The regulation of transcription is controlled by the level of the product of the pathway, tryptophan, through two mechanisms, the control of transcription initiation by *trp* repressor–operator interaction, with tryptophan acting as corepressor, and the control of transcription termination by the *trp* attenuator region. The nucleotide sequence of the promoter region determines the affinity for the RNA polymerase, and thus the rate of transcription. There is no CRP site in the *trp* operon.

The his, leu, phe, and thr Operons

Transcription of these operons and the synthesis of these enzymes is affected by levels of each amino acid and its uncharged tRNA through transcription termination at the attenuator region. The nucleotide sequence of the promoter region determines the affinity for the RNA polymerase, and thus the rate of transcription. All lack an operator region and a CRP site. (See Table 20.1 for information on these operons.)

20.10 REGULATION OF TRANSCRIPTION IN MAMMALS

There are many different types of mammalian cells, each containing different amounts and types of proteins at different times; in addition, it is estimated that only a fraction of the total DNA is transcribed in a particular mammalian cell. Thus it seems clear that transcription is regulated in mammals. Yet the underlying mechanism is unknown. One basic difficulty of gathering data in mammals and other animals is the slowness and difficulty of genetic analysis in animals as contrasted with bacteria. *E. coli* is a relatively simple and hardy cell, easily cultured and manipulated, with a fairly rigid cell wall. It has a generation time of about 30 min, and it is relatively easy to generate, select, and analyze mutants. By contrast, cell culture and genetic analysis in mammals is slow and difficult. The cells divide slowly, are easily damaged, and contain complex subcellular structures not found in bacteria (nuclei, chromosomes, mitochondria, endoplasmic reticulum, golgi apparatus, etc.). These differences account for much of the difficulty of obtaining data to clarify the control of transcription in mammals.

Nonetheless, it seems very likely that promoter regions exist in mammalian DNA; indeed as shown in Figure 20.5*a*, there are nucleotide sequences in SV40, a mammalian virus, that are remarkably similar to bacterial promoters. There are also mammalian genes that contain the sequence TATAA about 31 nucleotides upstream from the mRNA cap site (Figure 20.18); this sequence is thought to be a possible recognition site for RNA polymerase II, though data on this point are almost entirely lacking. It will be remembered (Chapter 18) that mammalian cells contain at least three types of RNA polymerases (type I for rRNA; type II for HnRNA and mRNA; and type III for low molecular weight RNA), and if the bacterial model applies, we can expect at least three or more homologous sets of promoter sequences. To date, however, no promoter mutants have been reported in mammals, and no putative promoter region has been shown to be protected from the action of DNase by RNA polymerase. Furthermore, no unequivocal mutant in a CRP site, operator region, attenuator region, or a repressor gene has been reported in mammals. A few mutants exist that might be of these types, but crucial data of the type that exists in bacteria that would allow clear-cut designation of mechanism is lacking (see Clin. Corr. 20.1).

Lacking the ability to select and analyze mutants has spurred researchers working on mammals and other animals to utilize other

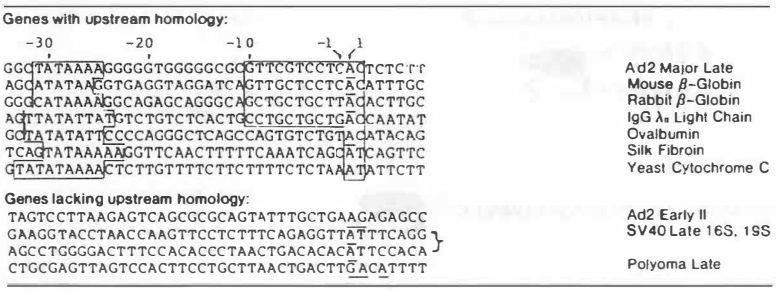


Figure 20.18

Nucleotide sequences of mammalian DNA preceding mRNA cap sites. *The nucleotide corresponding to the capped 5′ terminus of mRNA is underlined. The region of TATAAA homology around region −31 is indicated.*

Reproduced with permission from C. C. Baker, G. Herisse, G. Courtois, F. Galibut, and E. Ziff, *Cell*, 18:569, 1979. Copyrighted M.I.T.

techniques available to them. The analysis of the effect of steroid hormones on mammalian cells has been particularly revealing. As discussed in Chapter 15, steroid hormones increase anabolic processes in target cells by an increase in transcription of selected genes followed by an increase in protein synthesis. The details of how this comes about is not clear, but an overall picture is emerging (Figure 20.19). After entering a cell, steroids react with receptor molecules located in the cytoplasm; the receptor molecules are protein dimers containing A and B subunits. The hormone–receptor complex migrates into the nucleus, where the B subunit binds to specific acceptor sites on the chromosomes, presumably to the nonhistone chromosomal protein complexed with the DNA proximal to the gene whose transcription is to be affected. After the dimer is bound to the acceptor sites, the A subunit is released and binds to the DNA. This produces a change in the DNA so that transcription occurs. Thus this theory postulates that the hormone combined with the receptor protein exerts positive control of gene transcription. The model is thought to apply to the action of aldosterone, androgens, estrogen, glucocorticoids, and progesterone (Clin. Corr. 20.2). The similarity of this model to the interactions of cAMP, CRP, and the CRP site in *E. coli* is striking.

This model postulates an important role for the nonhistone (acidic) chromosomal proteins and no role for histones in controlling gene transcription in mammals, even though both are present in the chromosome. There are many reasons for this. Histones are few in number (only five major types normally exist in animal cells) and are

CLIN. CORR. 20.1 β^+ AND β^0-THALASSEMIAS

It is thought that the defect in some patients with certain types of thalassemias (the so-called β^+- and β^0-thalassemias) are due to either decreased transcription of the β-globin structural gene or abnormal processing of the mRNA. Using radioactive cDNA probes, it is possible to show that there are normal amounts of the gene coding for β-globin in these patients, but that there is deficiency of the mRNA of β-globin. Comparison of the nucleotide sequence of the β-globin gene and flanking regions in normal humans and in these patients might prove to be revealing.

CLIN. CORR. 20.2 MALE PSEUDOHERMAPHRODITISM

Male pseudohermaphroditism is a condition characterized by a failure of the development of male secondary sex characteristics. Experiments show that the target organs do not respond to androgens. This condition is a result of a defect in the androgen receptor proteins.

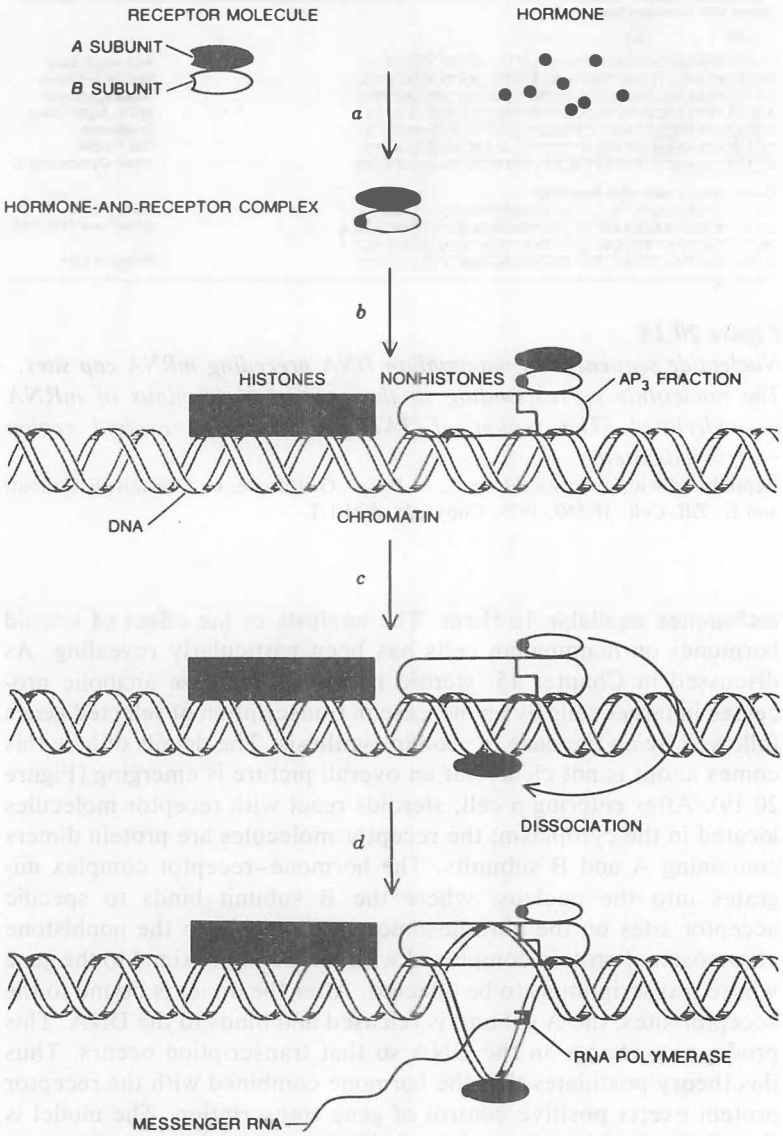


Figure 20.19
Mechanism of gene activation by the steroid hormones.
(a) A receptor molecule binds to two molecules of hormone. (b) This combination enters the nucleus and binds to the AP_3 fraction of the nonhistone chromosomal proteins of specific genes. The binding takes place between the nonhistone protein and the B subunit of the receptor molecule. (c) The subunits then dissociate, and the A subunit interacts with the DNA enabling a molecule of RNA polymerase to occupy an initiation site on the DNA. (d) A segment of DNA is then transcribed producing a strand of mRNA that serves as a template for the synthesis of a protein.

not distributed in amounts or in a way that would suggest they are controlling gene transcription. The more acidic nonhistone proteins, on the other hand, are diverse, are present in greater amounts in cells that are active in RNA synthesis, and exhibit tissue specificity and DNA-binding specificity. In addition, the nonhistone chromosomal proteins have been shown to stimulate RNA synthesis in vitro.

An interesting mutant in mice exists that might be affecting transcription or mRNA processing. Figure 20.20 shows the induction of β-glucuronidase in the kidney of female mice of two strains and the F_1 hybrid injected with dihydrotestosterone. The difference observed between the two strains of mice is not seen with two other enzymes also induced at the same time, alcohol dehydrogenase and arginase (Figure 20.21). The difference in inducibility of β-glucuronidase in these two mice strains is due to a difference in a single gene at or near the structural gene for β-glucuronidase on chromosome 5 of the mouse. Biochemical analysis showed that there are no obvious structural differences in the β-glucuronidase but that the rate of synthesis of this enzyme was shown to be different in the two strains. Of great interest is the discovery that frog oocytes synthesized catalytically active mouse β-glucuronidase after injection of mRNA extracted from induced mouse kidney; no β-glucuronidase is detected in frog oocytes injected with mRNA from uninduced kidneys. Using this assay system, it was possible to show a large difference in the activities of mRNA from the two mouse strains. Thus the difference in inducibility of β-glucuronidase with androgens between the kidneys of two mouse strains is either due to differences in the rate of transcription of the β-glucuronidase structural gene or in the processing of the mRNA after it is formed. Further work is required to distinguish between these possibilities; the nucleotide sequence of these genes may be quite revealing.

It is unfortunate that studies on the control of transcription in mammals have not kept pace with the discoveries in bacteria. For the reasons mentioned in the beginning of this section, this lag is bound to continue. However, the use of genetic engineering techniques to clone mammalian genes has already revealed a wealth of information about the nucleotide sequences of mammalian genes and their flanking regions, and it seems certain that additional clarification of the control of transcription will be revealed. Of particular interest will be how much of what is known about bacterial systems is applicable to mammals.

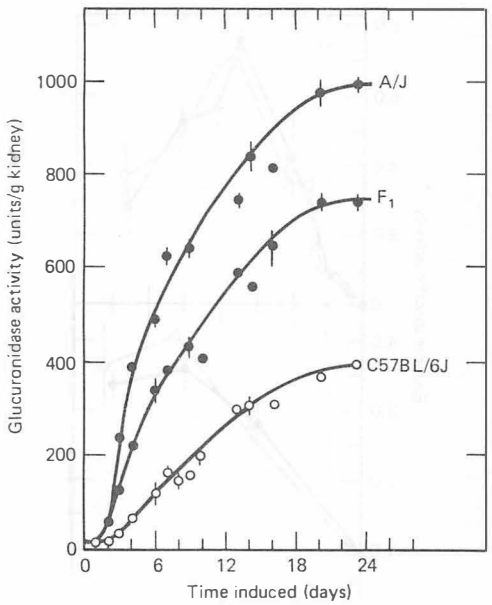


Figure 20.20
Graph showing differences in the levels of induction of kidney glucuronidase in high (A/J) and low (C57BL/6J) inbred strains and in their F_1 progeny.
Females of all three strains were injected with 10 mg dihydrotestosterone on day zero and on day 1, and with 5 mg every 3 days thereafter. The value plotted at each time point is the mean for two mice. The range is represented by the vertical bar. F_1 mice were the offspring of a cross between C57BL/6J females and A/J males. (●), A/J; (◑), F_1; (○), C57BL/6J.
Reproduced with permission from R. T. Swank, K. Paigen, and R. E. Ganschow, *J. Mol. Biol.*, 81:225, 1973. Copyright by Academic Press Inc. (London) Ltd.

20.11 THE BIOCHEMISTRY OF GENETIC ENGINEERING

An important aspect of genetic engineering is the transfer of genes from one organism to another so that the DNA of the transferred gene is replicated in the host organism; if the DNA attaches to or contains a regulatory region containing a promoter region, it may be

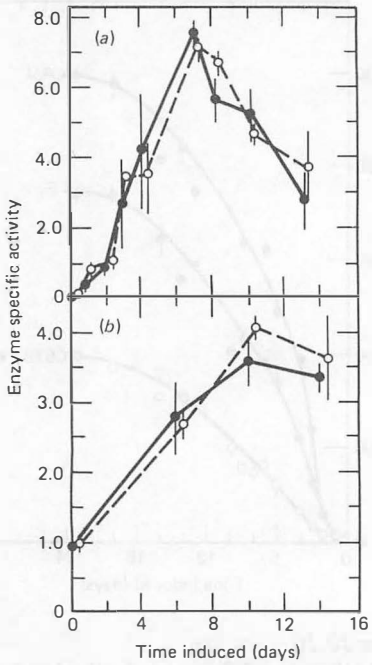


Figure 20.21

(a) Graphs showing no differences in the induction of kidney alcohol dehydrogenase by dihydrotestosterone in strains A/J and C57BL/6J. The injection schedule is described in the legend to Figure 20.20. Two animals of each strain were assayed for alcohol dehydrogenase at each time point. The value plotted at each time point is the mean activity of two mice. The range is represented by the vertical bar. Specific activity is expressed as nanomoles NADH oxidized per hour per milligram kidney. (●), A/J; (○), C57BL/6J. (b) Graphs showing no differences in the induction of kidney arginase by dihydrotestosterone in strains A/J and C57BL/6J. Protocol as in (a). Specific activity is expressed as micromoles urea formed per hour per milligram kidney. (●), A/J; (○), C57BL/6J.

Redrawn with permission from R. T. Swank, K. Paigen, and R. E. Ganshono, *J. Mol. Biol.*, 81:225, 1973. Copyright by Academic Press Inc. (London) Ltd.

transcribed and translated; if the codon reading frame is correct, a functional protein may be produced in a cell that never produced this protein before. Most life forms accomplish this task routinely as part of the sexual cycle. Another aspect of genetic engineering is the ability to alter the nucleotide sequence of known genes (*site-specific mutagenesis*). The recent acquisition of knowledge concerning enzyme and nucleic acid technology now gives scientists the ability to accomplish both types of gene manipulation in the laboratory by methods not thought possible a decade ago. This is one of the achievements of molecular biology. The availability of techniques to clone genes makes available a research and clinical tool of immense potential. Indeed, the ability to obtain sufficient quantities of DNA for the determination of nucleotide sequence has led to many of the facts and ideas presented in this chapter. In this section we discuss some aspects of the technology of genetic engineering, as well as exciting potential clinical applications.

A General Method for Transfer of DNA into E. coli

The cloning of many mammalian genes in *E. coli* has already been accomplished; Table 20.3 provides a partial list. The methodology involves obtaining the DNA of the desired gene; placing the DNA into a *vector* or vehicle capable of transporting the gene and maintaining it inside an *E. coli* cell; inserting the vector containing the gene into *E. coli;* detecting and selecting the *E. coli* cells containing

Table 20.3 A Partial List of Genes Coding for Mammalian Proteins Cloned in *E. coli*

MOUSE	RAT
Dihydrofolate reductase	Insulin
α- and β-Globin	Growth hormone
K-Light chain immunoglobin	Muscle actin
α-Heavy chain immunoglobin	Myosin light chain
γ-1 Heavy chain immunoglobin	Serum albumin
α- and β-Thyrotropin	
RABBIT	HUMAN
α- and β-Globin	Chorionic somatomammotropin
BOVINE	Growth hormone
Parathyroid hormone	Interferon
	Insulin
	α-, β-, γ-, ε-Globin
	Somatostatin

the gene attached to the vector; isolating the cloned gene from the *E. coli* and from the DNA of the vector, confirming its identity by determining its nucleotide sequence; and determining whether the gene is functional in *E. coli* by detecting the protein for which it codes. The technology for accomplishing this is diverse and often ingenious. Each gene presents its own challenges and opportunities for exploitation by creative scientists.

Figure 20.22 shows one general method for introducing foreign DNA into a plasmid for cloning in *E. coli*. The plasmid is a small circular DNA molecule that replicates independently of the DNA of the host cell; in order to do this it contains the nucleotide sequences required for self-replication. In addition, the plasmids used in genetic engineering contain nucleotide sequences that code for proteins that confer resistance to antibiotics onto the host; cells carrying the plasmid can be detected and selected for on media containing the antibiotic. This makes the plasmid an excellent vector for the genes to be cloned. As shown in Figure 20.22, the DNA of the plasmid and the foreign DNA is cleaved by a restriction endonuclease, which leaves the same unpaired ends on both molecules. The cleaved plasmid and foreign DNA fragments are brought together under conditions promoting the formation of hydrogen bonds between the ends, a process called *annealing*. The breaks in the nucleotide sequence are repaired with DNA ligase. This enzyme repairs breaks of this type in one strand of the DNA by catalyzing the synthesis of a bond between the phosphate of one nucleotide and the deoxyribose of the adjacent nucleotide. The plasmid containing the foreign DNA is introduced into *E. coli* by a process called *transformation*. The *E. coli* are plated out on media containing the antibiotic(s) to which the plasmid confers resistance. Thus only those *E. coli* cells containing the plasmid can live and replicate. Further screening and selection of *E. coli* containing the plasmid plus the foreign DNA is carried out by special techniques. The cloned, foreign DNA can be recovered for sequence analysis and other purposes first by the isolation of total *E. coli* DNA and the separation of the plasmid DNA from the *E. coli* DNA by centrifugation in cesium chloride and ethidium bromide, and then by cleaving the boundaries between the plasmid DNA and the foreign DNA with the same restriction endonuclease used in the beginning of the process. The smaller, foreign DNA fragment can be easily separated from the much larger plasmid DNA by gel electrophoresis or column chromatography.

Cloning the Mammalian Gene That Codes for Insulin

The biosynthesis of animal peptides and proteins in *E. coli* have been reported. This includes human γ and β (hemo)globins, chicken oval-

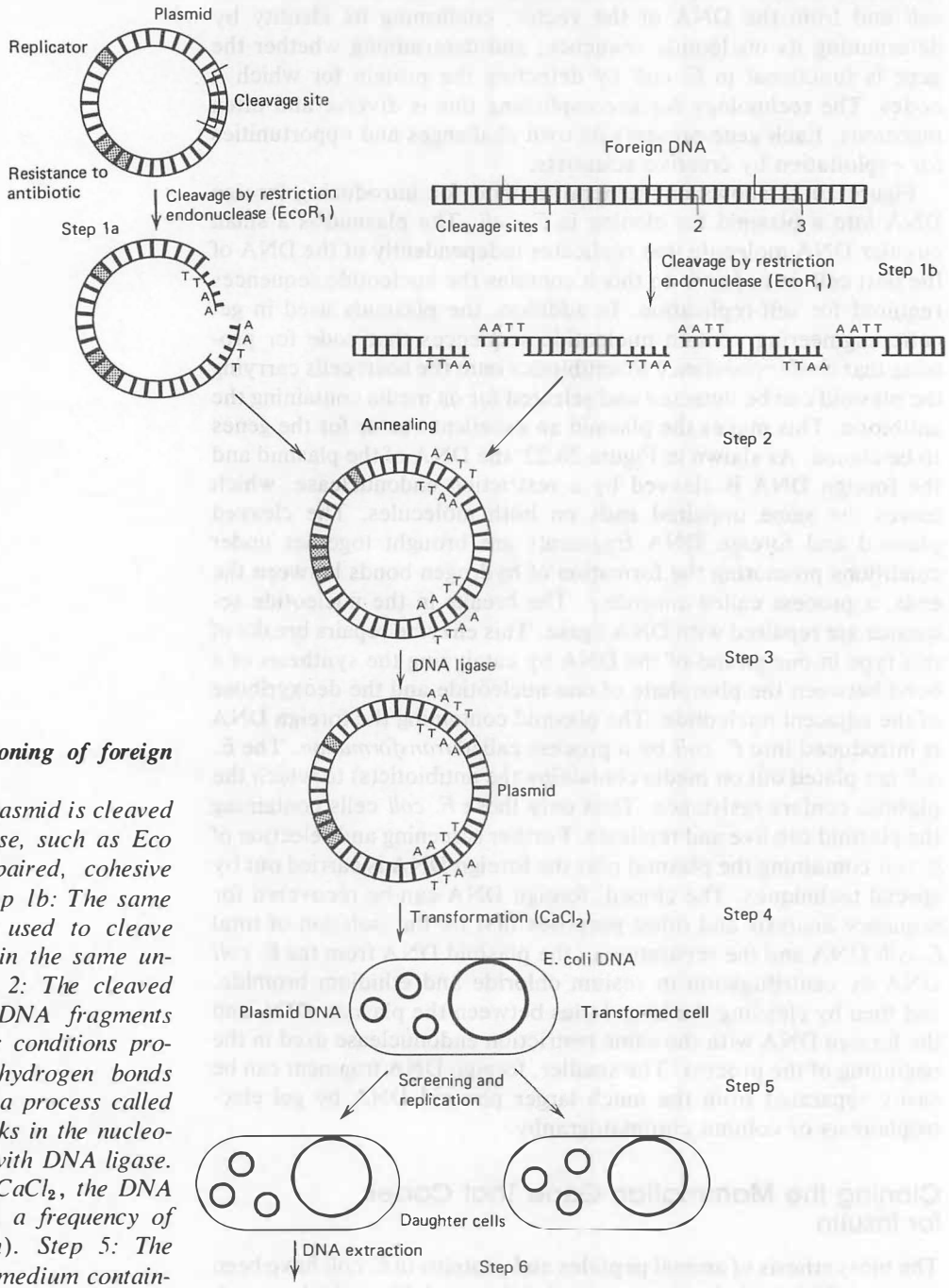
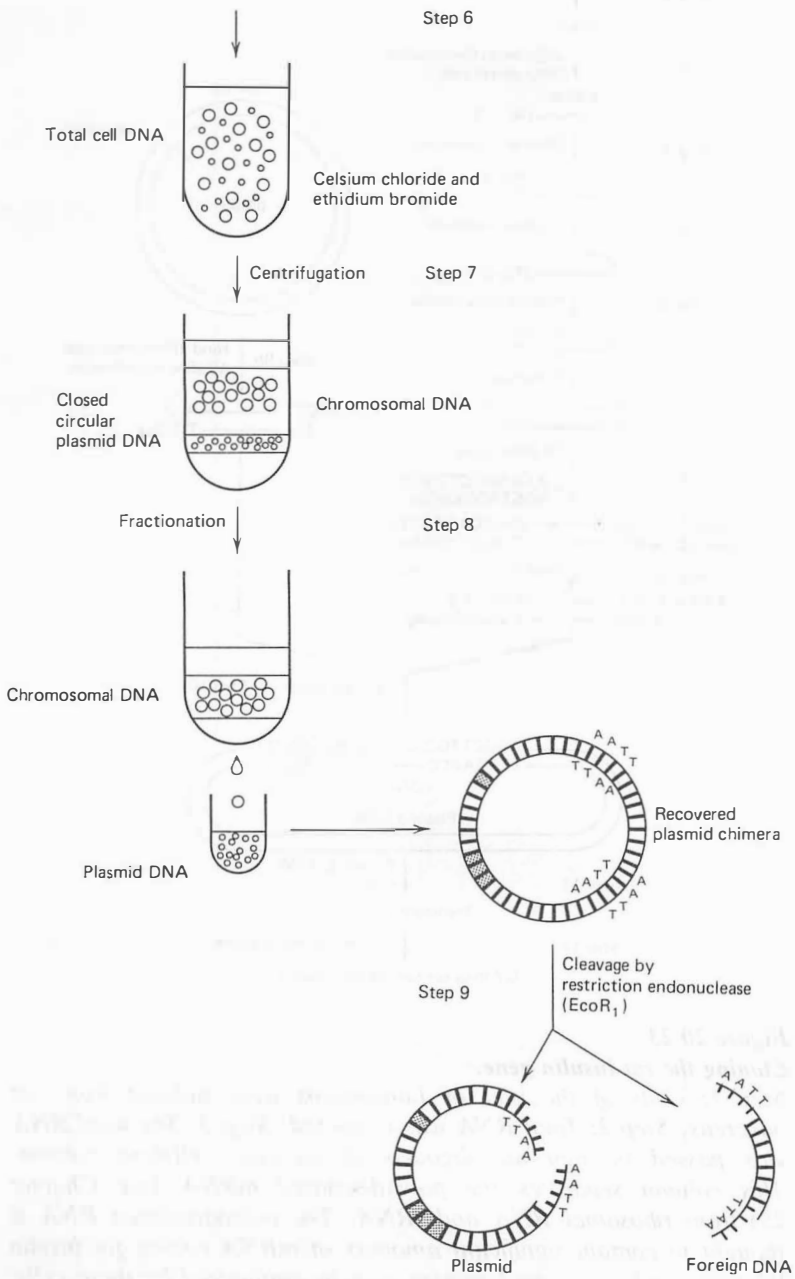


Figure 20.22
***General approach to the cloning of foreign DNA in* E. coli.**
Step 1a: The DNA of the plasmid is cleaved by a restriction endonuclease, such as Eco R_1, which leaves the unpaired, cohesive ends, TTAA and AATT. Step 1b: The same restriction endonuclease is used to cleave the foreign DNA resulting in the same unpaired,cohesive ends. Step 2: The cleaved plasmid and the foreign DNA fragments are brought together under conditions promoting the formation of hydrogen bonds between the cohesive ends, a process called annealing. Step 3: The breaks in the nucleotide sequence are repaired with DNA ligase. Step 4: In the presence of $CaCl_2$, the DNA enters the E. coli cell with a frequency of about 10^{-5} (transformation). Step 5: The E. coli are planted out on a medium contain-



ing the antibiotic(s) to which the plasmid confers resistance and only those E. coli *cells containing the plasmid can live and replicate. Step 6: To separate the cloned, foreign DNA from E. coli the total DNA is first isolated. Step 7: The plasmid DNA is separated from* E. coli *DNA by centrifugation in cesium chloride and ethidium bromide. Step 8: The plasmid DNA is collected from the centrifuge tube. Step 9: The nucleotide boundaries between the plasmid DNA and the foreign DNA in the plasmid are cleaved with the same restriction endonuclease used to cleave the DNA in steps 1a and 1b, above. The smaller, foreign DNA fragment can be easily separated from the much larger plasmid DNA by gel electrophoresis or column chromatography.*

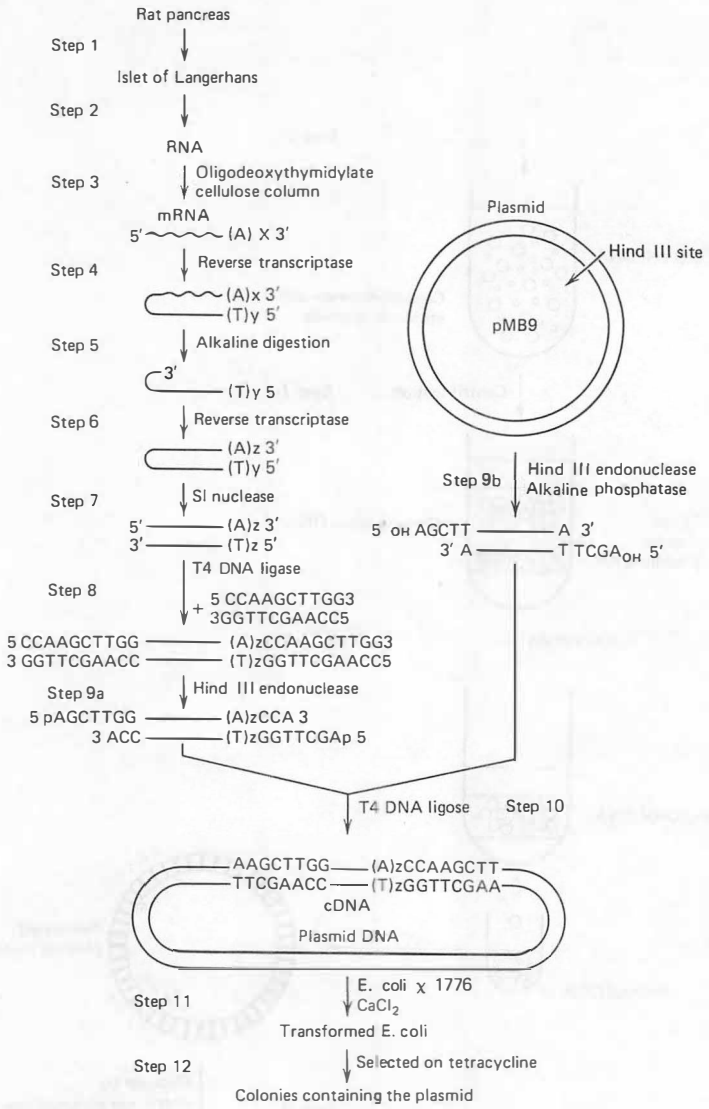


Figure 20.23
Cloning the rat insulin gene.
Step 1: Cells of the islet of Langerhans were isolated from rat pancreas. Step 2: Total RNA was extracted. Step 3: The total RNA was passed through an oligodeoxythymidylate cellulose column. This column separates the polyadenylated mRNA (see Chapter 18) from ribosomal RNA and tRNA. The polyadenylated RNA is thought to contain significant amounts of mRNA coding for insulin because insulin is a predominant protein synthesized by these cells. Step 4: The mRNA was transcribed into cDNA using reverse transcriptase, the viral enzyme that transcribes DNA on RNA templates

bumin, human interferon, human insulin, and mouse β-endorphin. Three variations of the general method described above have been used to clone the mammalian gene that codes for insulin. Two of the methods started with rat insulin *cDNA* (Figures 20.23 and 20.24), that is, DNA transcribed from the mRNA of the desired gene by reverse transcriptase, the viral enzyme that synthesizes DNA on RNA templates (see Chapter 18). The third method (Figure 20.26) started with chemically synthesized DNA, the nucleotide sequence of which was based on the amino acid sequence of the A and B

(see Chapter 18). Step 5: The mRNA was hydrolyzed with alkali. Step 6: The single-stranded cDNA was converted into a double-stranded DNA helix by using reverse transcriptase. Step 7: The resulting double helix contains a single-stranded hairpin loop that was removed by S1 nuclease, an enzyme that hydrolyzes single-stranded DNA. Step 8: The cDNA is now a double helix with an unknown number of A · T base pairs at one end. In order to produce cohesive ends for the introduction of this cDNA into a plasmid, the chemically synthesized decanucleotide shown in the figure was attached to both ends using DNA ligase. This decanucleotide contains the palindromic symmetry recognized by the Hind III restriction nuclease. Step 9a: Treatment with the Hind III restriction nuclease produces a cDNA molecule with Hind III cohesive ends. Step 9b: The plasmid pMB9, which contains tetracycline resistance and other factors, was cleaved with Hind III restriction endonuclease and exposed to bacterial alkaline phosphatase, an enzyme that removes the phosphates from the cleaved 5'-terminal ends of the plasmids at the Hind III site. This prevents the cleaved plasmid from reforming without the insertion of the cDNA and reduces the number of plasmids that will reform into circles without insertion of the insulin gene. Step 10: The linear plasmid and the insulin gene are mixed, and annealing into circles containing DNA from the plasmid and the insulin gene is allowed to take place. DNA ligase is used to repair the breaks. Step 11: This mixture was used to transform E. coli *strain x1776. Step 12: Strains of* E. coli *that took up the plasmid were selected by their ability to grow on tetracycline.*

The insulin gene was detected by isolating the plasmid DNA and determining the nucleotide sequence of a small DNA fragment released from the plasmid DNA by Hind III restriction endonuclease; the observed nucleotide sequences corresponded to those expected based on the known amino acid sequence of rat insulin.

Redrawn with modifications and permission from A. Ullrich, et al., *Science*, 196:1313, 1977. Copyright 1977 by the American Association for the Advancement of Science.

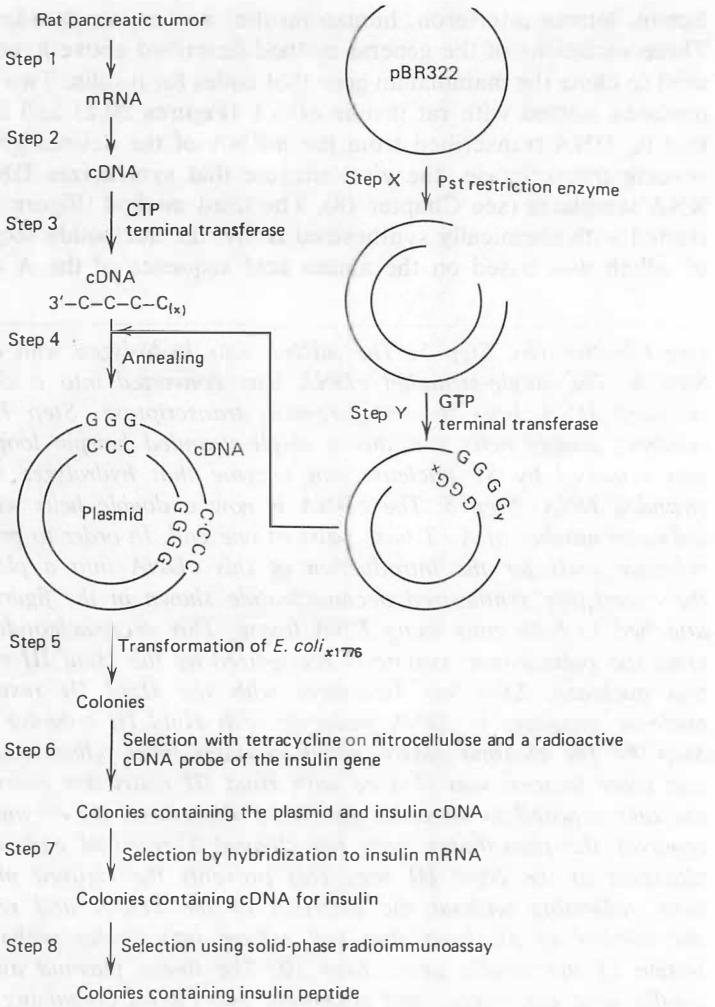


Figure 20.24
Cloning the structural gene coding for rat insulin.
Step 1: mRNA was isolated from a rat pancreatic tumor. Step 2: cDNA was transcribed from the mRNA. Step 3: To form cohesive ends, a string of dC nucleotides were added to the 3′ ends of cDNA using CTP plus terminal transferase, an enzyme that will add specific nucleotides to the 3′ end of DNA. Step X: The plasmid pBR322 was cleaved with Pst restriction endonuclease. Step Y: dG nucleotides were added to the 3′ ends using GTP plus terminal transferase. Thus the dC at the 3′ end of the cDNA and the dG at the 3′ end of the cleaved plasmid constitute the unpaired cohesive ends for annealing. Step 4: The plasmid DNA and the cDNA were mixed and annealed. Step 5: The presumed hybrid DNA was used to transform E. coli *strain x1176. In this case the breaks were not*

chains of human insulin. The cohesive ends for annealing between the insulin DNA and the plasmid DNA was different in each case. A chemically synthesized decanucleotide containing the palindromic symmetry recognized by the *Hind* III restriction nuclease was used in one (Figure 20.22), while the other used the enzyme, terminal

repaired with DNA ligase; the DNA ligases in the E. coli *cells were allowed to do that. Step 6: To select for cells that contained the recombinant plasmid, the* E. coli *that had been exposed to the plasmid were spread on sterile nitrocellulose filters lying on agar medium containing tetracycline in petri dishes. The plates were incubated for 2 days to allow those* E. coli *that were resistant to tetracycline, and thus contained the plasmid, to form colonies. A portion of each colony was saved by pressing sterile filter paper over them; this filter paper was placed on fresh media and saved. The colonies on the nitrocellulose filter containing the insulin gene were detected by lysing the cells and denaturing and fixing the DNA to the nitrocellulose filter, in situ. The DNA was exposed to a radioactive probe consisting of the cDNA of the insulin gene cloned in Figure 20.22. The radioactive probe will hybridize and anneal to the DNA fixed to the filter that contains nucleotide sequences of the insulin gene. This combination can be detected by autoradiography. Step 7: Colonies that gave this positive result were further screened for the ability of the DNA taken from the plasmid to hybridize with insulin mRNA and interfere with the ability of this mRNA to code for insulin in a cell-free protein-synthesizing system. One clone passed this test. The insulin gene was removed from the plasmid DNA in this clone using Pst restriction endonuclease and its nucleotide sequence verified the presence of the rat insulin gene (Figure 20.25). By using this cloned insulin gene as a radioactive probe it was possible to isolate 48 strains of* E. coli *that contained parts of a cloned insulin gene. Step 8: These strains were screened for their ability to synthesize insulin by using a solid-phase radioimmunoassay. The* E. coli *cells were lysed and placed in contact with a polyvinyl sheet to which antibody to rat insulin was attached. Insulin, if present, combines with the antibody; this combination, in turn will combine with* ^{125}I*-labeled antibody to insulin that is added last. The radioactivity can be detected by autoradiography. One strain gave positive results in this test; indeed it was estimated that this strain produced 100 molecules of rat insulin per cell. This screening method has important implications and uses in cloning other genes.*

Adapted with permission from L. Villa-Komaroff, A. Efstratiadis, S. Broome, et al., *Proc. Nat. Acad. Sci. USA*, 75:3727, 1978.

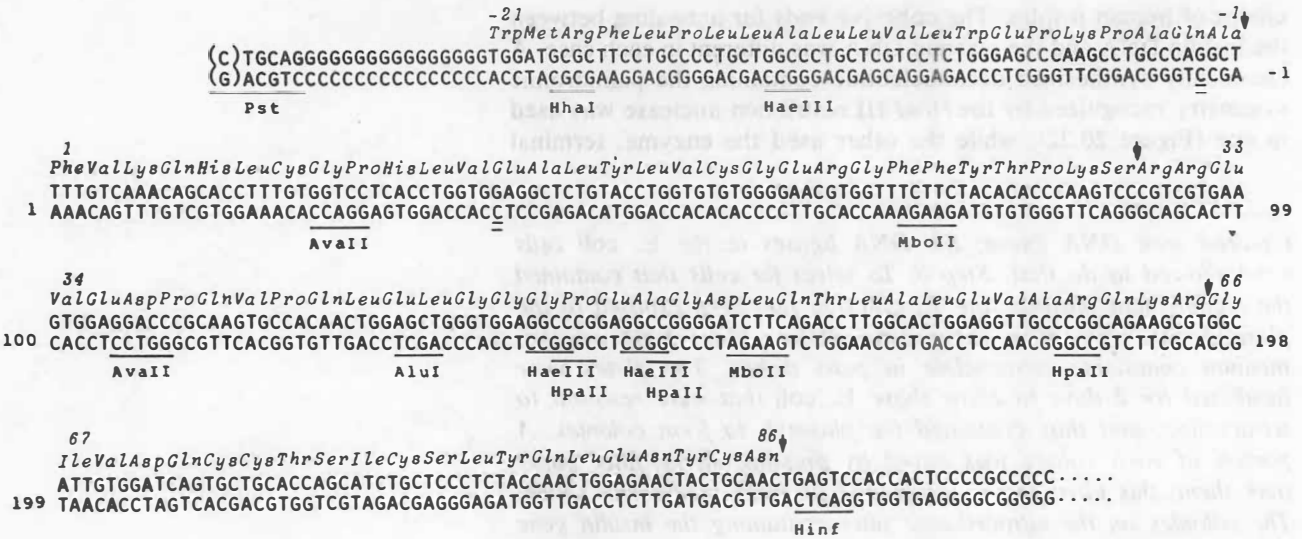


Figure 20.25

Nucleotide sequence and amino acid codons of a clone containing a portion of the rat insulin gene.

Nucleotides are numbered beginning with the first base in the sequence encoding proinsulin. Amino acids are numbered beginning with the first amino acid of proinsulin, while the last amino acid of leader sequence is numbered as −1. Restriction endonuclease cleavage sites experimentally verified are underlined and identified. The arrows indicate, in order, the ends of the leader sequence and the peptides B, C, and A. Two nucleotides indicated by double underlining are uncertain.

Reproduced with permission from L. Villa-Komaroff, A. Efstratiadis, S. Broome, et al., *Proc. Natl. Acad. Sci. USA*, 75:3727, 1978.

transferase, to add chains of poly(C) to the 3′ ends of the cDNA and chains of poly(G) to the 3′ ends of the cleaved plasmid (Figure 20.24). The methods used to select the colonies containing the insulin DNA, the insulin mRNA, and the insulin peptides are explained in the legends to Figures 20.24 and 20.26. All three studies used *E. coli* strain *x* 1776. This strain is specially chosen because it does not survive in the intestines of mammals where *E. coli* normally thrives, since it requires diaminopimelic acid, threonine, methionine, biotin, and thymine for growth.

Of interest is the biosynthesis of insulin fused to other proteins designed to expedite its isolation. For example, in the procedure described in Figure 20.23, the *Pst* site cleaved in the plasmid was at

the end of a plasmid gene coding for penicillinase just before the signals for termination of transcription and translation. Thus the rat insulin protein was fused to *E. coli* penicillinase, an enzyme that is normally found in the periplasm, the space between the *E. coli* cell membrane and the cell wall. The fused protein was indeed located in the periplasmic space, and could be removed by osmotic shock, a procedure that removes the cell wall, but does not break open the cell. Thus the rat insulin protein was relatively free of other *E. coli* cellular proteins. As shown in Figure 20.26, the chemically synthesized DNA coding for each insulin chain was inserted into a plasmid containing a portion of the *lac* operon, such that the strains of *E. coli* that carried the recombinant plasmid produced the insulin chain fused to β-galactosidase. This fused protein constituted about 20% of the total cellular protein, and because it was insoluble was easily obtained in the pellet after low-speed centrifugation of crude extracts. The A and B chains of insulin were isolated, purified, and chemically attached to one another to produce an insulin molecule easily detectable in a radioimmunoassay sensitive to human insulin. The final yield was estimated at about 10 mg of human insulin from 24 g of wet *E. coli* cells; each *E. coli* cell contained about 10,000 molecules of the fused protein.

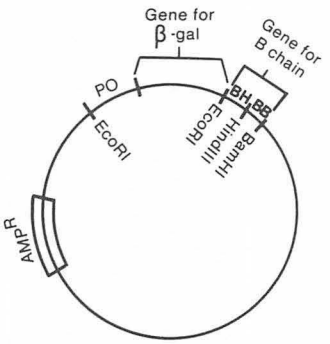


Figure 20.26
Schematic representation of the plasmid containing the structural genes for the β chain of human insulin.
AMP^R (resistance for ampicillin; structural gene for penicillinase); PO (promoter and operator region of the lac operon); Gene for β-gal (structural gene for β-galactosidase); gene for β-chain (structural gene for the β chain of insulin); BH, BB (two portions of the gene coding for the β chain of insulin); Eco-RI, HindIII, and BamHI—sites for cleavage by specific restriction endonucleases.
Reproduced with permission from D. Goeddel, D. G. Kleid, F. Bolivar, et al., *Proc. Nat. Acad. Sci. USA*, 76: 106, 1979.

Potential Clinical Applications of Genetic Engineering

Using techniques similar to those described above, genetic engineers have been able to clone many genes from a variety of organisms, including bacteria, yeast, fungi, higher plants, and animals. The list of all the genes cloned in *E. coli* is too long to present here. The clinically relevant genes that have been cloned include the mammalian genes shown in Table 20.3. There are two major clinical goals of genetic engineering. The first is to be able to produce clinically useful proteins inexpensively and conveniently in *E. coli;* such proteins might include insulin, the blood-clotting factors, protein hormones, interferon, and so on. This goal seems to be realizable in the not too distant future once certain technical problems are resolved. A second clinical goal is to be able to use cloned normal human genes to replace the defective genes of patients with genetic diseases such as diabetes, sickle cell anemia, and hemophilia. To do this requires much more than producing large amounts of cloned DNA. Technology has to be developed to introduce the DNA into the appropriate cells of the patient so that these cells are transformed. In addition, it is important that the transcription of the introduced gene is regulated normally in the cell, since inappropriate transcription could be detrimental to the patient.

20.12 TRANSFORMATION OF MAMMALIAN CELLS

Various approaches to the introduction of cloned DNA into mammalian cells are being studied, and a number of methods seem potentially useful. These include using viruses as vectors, DNA-mediated gene transfer, and using DNA in phospholipid vesicles (liposomes). Genes have also been introduced into mammalian cells as part of metaphase chromosomes.

Using Viral Vectors

The use of viruses as vectors seems feasible, since the virus contains the signals necessary for the DNA to enter mammalian cells and replicate. Using methodology similar to the type used to insert foreign DNA into bacterial plasmids, it has been possible to insert mammalian DNA into viruses, to infect mammalian tissue culture cells with the virus, and to detect the foreign DNA and, in some cases, detect mRNA and the protein coded for by the foreign DNA. For example, cloned rabbit β-globin cDNA was inserted into the circular DNA of the SV40 virus in place of a late developmental gene coding for the major capsid protein of the virus. In another study, the DNA of a mutant mouse β^{maj}-globin gene, complete with intervening and flanking sequences, was inserted into SV40 in place of the late developmental genes. Monkey tissue culture cells infected with either of these viroids synthesized mouse or rabbit globin mRNA detected by hybridization to a radioactive probe. Apparantly mRNA processing (splicing, polyadenylation, etc.) occurred normally because the cells produced mouse β^{maj}-globin or rabbit globin peptide identified by radioimmunoassay, two-dimensional gel electrophoresis, and tryptic peptide mapping. Thus it is clear that in these cases rabbit and mouse β-globin DNA was transcribed, and that rabbit and mouse nucleotide signals for mRNA processing and translation are operational in monkey cells.

DNA-Mediated Transformation

It has been reported that DNA containing the gene coding for thymidine kinase (tk^+) from a variety of sources (herpes simplex virus, human cells, chinese hamster cells, etc.) can be taken up by endocytosis of microprecipitates of calcium–DNA complexes and expressed in tk^- mammalian tissue culture cells; the frequency of transformed from tk^- to tk^+ cells is about 10^{-5}. The tk^+ cells are

selected for in a special medium containing hypothanine, adenine, and thymidine (called *HAT medium*); tk^- cells do not survive in this medium. The thymidine kinase produced corresponds to the type made by the donor of the DNA, not the host cell type. Similar results have been obtained using mammalian DNA containing the gene that codes for adenine phosphoribosyltransferase ($aprt^+$) exposed to $aprt^-$ tissue culture cells.

DNA-mediated transformation can also be carried out by two methods for genes for which no selection system exists: *cotransformation* and by *chemical linkage*. Cotransformation consists of mixing the DNA containing the tk^+ gene with the DNA coding for other genes. Tissue culture cells that are tk^- are exposed to this mixture of DNA, and the transformed tk^+ cells are selected for in HAT medium. The assumption made here is that cells that are *competent* to take up the tk^+ DNA will also take up other DNA as well. In one such study, DNA coding for the rabbit β-globin gene was used with tk^+ DNA to cotransform mouse tk^- cells. Many of the tk^+ transformed colonies also contained the DNA and mRNA coding for rabbit β-globin, as shown by hybridization to a radioactive cDNA probe.

Another way to extend this approach to genes for which no selection system exists is by chemical linkage to the tk^+ gene. In one such study, cloned tk^+ DNA was attached to cloned rabbit β-globin DNA through the use of restriction endonucleases and DNA ligase, as described above. This *concatenate* was used to transform tk^- mouse tissue; the tk^+ cells were selected on HAT medium. Almost all of the 27 surviving colonies selected for study contained the DNA for rabbit β-globin.

In another study the entire gene for chicken ovalbumin (plus the flanking regions) and three copies of the tk^+ gene were cloned in the same plasmid. The DNA from this plasmid was used to transform tk^- mouse tissue culture cells; the tk^+ cells were selected by their ability to grow on HAT medium. All transformants containing the tk^+ gene also contained the ovalbumin gene and produced chicken ovalbumin protein detected by radioimmunoassay. The various transformed colonies contained from 1,000 to 100,000 molecules per cell of this protein. From these data, it was concluded that this chicken gene can be transcribed, the mRNA can be processed correctly, and translation can occur appropriately in mouse cells. In addition, because of the way the ovalbumin gene was oriented in the DNA, it was thought that the promoter of the ovalbumin gene was used to initiate transcription. If this is true, then transcription might be affected by estrogen, the hormone that activates transcription of the ovalbumin gene in the oviduct of the chicken, that is, if the appropriate receptors are available in these tissue culture cells.

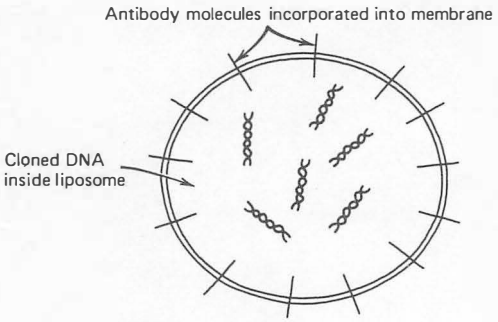


Figure 20.27
Schematic diagram showing a theoretical liposome vehicle of potential use in genetic engineering.

Chromosome-Mediated Transformation

Mammalian cells in tissue culture that are deficient in the enzyme hypoxanthine guanine phosphoribosyltransferase (HGPRTase) are unable to grow in HAT medium, and die. If these cells are exposed to metaphase chromosomes isolated from cells containing the gene coding for HGPRTase, about 10^{-6} cells grow on HAT media, and now contain HGPRTase of the type found in the cells that donated the chromosomes. If the metaphase chromosomes are enveloped in a liposome made from cholesterol and egg white, the frequency of surviving cells rises to about 10^{-5}. In addition, not only is the HGPRTase of the type from the donor cell, but other X-linked enzymes (such as, glucose 6-phosphate dehydrogenase and phosphoglycerate kinase) of the donor type are present in the surviving cells. It seems that the genes of the donor X-chromosome were incorporated as a unit into the genome of the recipient cell.

The Use of Phospholipid Vesicles

The use of a phospholipid vesicle (liposomes) to coat and protect the DNA from enzymatic hydrolysis may greatly enhance the potential of this technique. Such liposomes have been shown to protect bacterial plasmids from DNase and to increase the rate that genes in metaphase chromosomes are taken up and incorporated into the genome of mammalian cells. The usefulness of the liposome vector might be enhanced if antibodies against the surface of the appropriate target cells were incorporated into the surface, so the liposome would more readily adhere to the target cell (Figure 20.27). For example, one might imagine cloned human insulin genes in a liposome with antibodies to the β cells of the islet of Langerhans in the pancreas to treat diabetics, or cloned human globin genes in a liposome with antibodies against appropriate precursor cells to the red blood cells in the bone marrow to treat sickle cell anemia. This approach might be enhanced even more if one could remove the cells or tissues from the patient and expose these cells directly to the cloned DNA, naked, attached to a viral vector, or encapsulated in a liposome vehicle. Such a possibility exists for bone marrow cells (see Clin. Corr. 20.3) and for the β cells of the islet of Langerhans of the pancreas, both of which can be removed and placed back into mammals. We are still a long way from achieving this goal of genetic engineering, but there may well come a time when physicians will write prescriptions for gene therapy.

CLIN. CORR. 20.3 GENE TRANSFER IN INTACT ANIMALS

Gene transfer has been achieved in mice by exposing bone marrow cells to DNA isolated from tissue culture cells resistant to methotrexate. The resistance of these cells to this drug is due to the presence of

BIBLIOGRAPHY

Adhya, S., and Gottesman, M. Control of transcription termination. *Annu. Rev. Biochem.* 47:967, 1978.

Brown, D. D. Gene expression in eucaryotes. *Science* 211:667, 1981.

Cline, M. J., Stang, H., Mercola, K., et al: Gene transfer in intact animals. *Nature* 284:422, 1980.

Corden, J., Wasylyk, B., Buchwalder, A., et al: Promoter sequences of eukaryotic protein-coding genes. *Science* 209:1406, 1980.

Danchin, A., and Ullmann, A. The coordinate expression of polycistronic operons in bacteria. *Trends Biochem. Sci.* 5:51, 1980.

de Crombrugghe, B., and Pasten, I. Cyclic AMP. The cyclic AMP receptor protein, and their dual control of the galactose operon, in J. H. Miller and W. S. Resnikoff (eds.), *The operon.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1978, p. 303.

Gilbert, W., and Villa-Kamaroff, L. Useful proteins from recombinant bacteria. *Sci. Am.* 242:74, 1980.

Mulligan, R. C., and Berg, P. Expression of a bacterial gene in mammalian cells. *Science* 209:1422, 1980.

Nagata, S., Hideharu, T., Hall, A., et al: Synthesis in E. coli of a polypeptide with human leucocyte interferon activity. *Nautre* 284:316, 1980.

Pellicer, A., Robins, D., Wald, B., et al: Altering genotype and phenotype by DNA-mediated gene transfer. *Science* 209:1414, 1980.

Rosenberg, M., and Court, D. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 13:319, 1979.

high levels of dihydrofolate reductase, an enzyme involved in folic acid metabolism. The high levels of this enzyme are caused by amplification of the number of structural genes that code for this enzyme. DNA was isolated from a strain of mouse cells in tissue culture containing about 30 copies per cell of the structural gene for dihydrofolate reductase. By means of DNA-mediated transfer, this DNA was used to transform bone marrow cells taken from mice carrying a distinctive chromosomal marker. These cells, together with untreated cells lacking the chromosomal marker, were injected into immunologically compatible mice lacking the chromosome marker that had been irradiated to decrease haematopoiesis. After the injected bone marrow cells restored the formation of the blood cells, the mice were treated with methotrexate to inhibit cellular growth of nonresistant bone marrow cells and thus select for resistant cells. There was a clear increase in the percentage of bone marrow cells carrying the chromosomal marker, showing that they were transformed by the exposure to the DNA carrying the resistant genes. It was possible to remove these bone marrow cells and inject them into other mice, where they also grew well. Similar studies using the thymidine kinase (tk^+) gene from herpes simplex as transforming DNA were also successful.

These studies suggest that it may be possible to transfer genes for specific drug resistance to patients undergoing chemotherapy to enable them to tolerate higher doses of the anticancer drug, as well as make more feasible the idea of using genetic engineering to alleviate genetic disorders such as sickle cell anemia.

21

Metabolism of Individual Tissues

JOHN F. VAN PILSUM

21.1 SKELETAL MUSCLE

Structure

Skeletal muscle is composed of fibers that are striated in both the longitudinal and transverse directions. The longitudinal striations are the result of the fiber containing many myofibrils, and the transverse striations are due to the fact that the composition of the myofibril varies regularly along its length. A diagrammatic representation of skeletal muscle fibers and myofibrils is shown in Figure 21.1. Two main striations of the myofibril have been identified by light microscopy. The A band, or anisotropic band, is birefringent; it appears light in the dark field of a polarizing microscope because the light properties of the substance in this band are unequal in all directions. This property suggests a regular geometric arrangement of the molecules in the A band. The I, or isotropic band, is not birefringent, indicating that its light properties are equal in all directions, and it contains a substance of a more random arrangement. Other bands have been detected in the myofibril by staining procedures and light microscopy. A diagrammatic representation of the structure of the myofibril is shown in Figure 21.2. The Z band appears as a narrow dark line in the middle of the I band. The H zone appears in the center of the A band. The sarcomere is defined as the length of a myofibril bounded by two adjacent Z bands. In the center of the sarcomere at the H zone is an M line. The myofibril is composed of

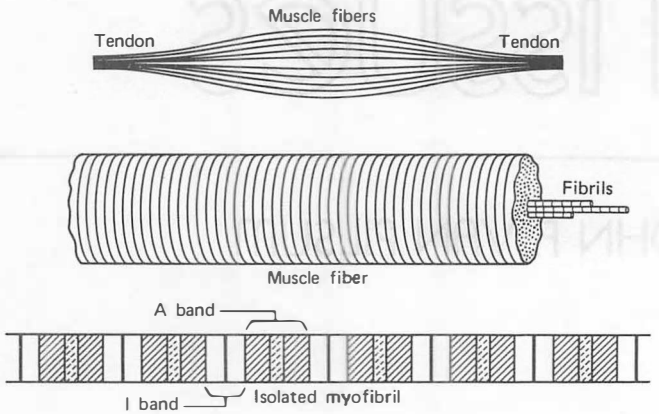


Figure 21.1
Diagrammatic representation of skeletal muscle fibers and myofibrils.

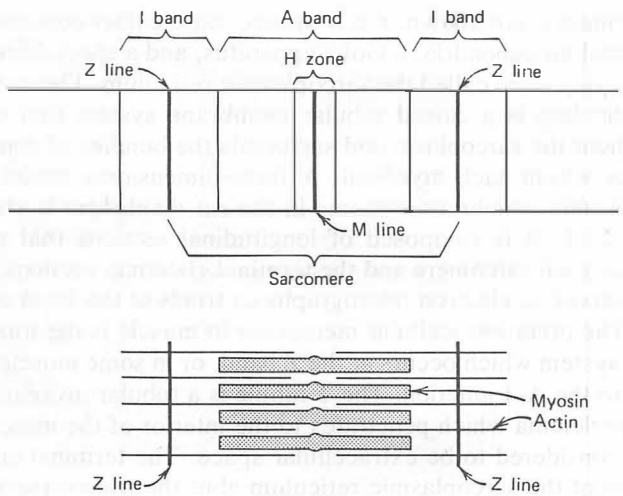


Figure 21.2
Diagrammatic representation of the structure of the myofibril.

thick and thin myofilaments, and the bands observed by light microscopy can be explained by the regular arrangement of the thick and thin myofilaments in the myofibril. The H and the I bands contain only thick or thin filaments, respectively, whereas the A band contains both thick and thin filaments. The Z line is a dense amorphous material to which the thin filaments are attached and the M line is an enlargement of the thick filaments. A regular spatial arrangement of the thick and thin filaments around each other is also known to occur. Each thick filament is surrounded by six thin filaments and each thin filament is surrounded by three thick filaments. The thick filaments are studded with projections that touch the thin filaments. The thick and thin filaments contain the proteins myosin and actin, respectively.

Each muscle fiber is covered with a sarcolemma, which has an outer layer of collagenous and reticular fibers that insert into the basement membrane. The sarcolemma is not a smooth membrane but has small indentations that act as a pinocytotic system for the transport of materials into the fibers. The myofibrils are imbedded in the cytoplasm of the muscle cell, called the sarcoplasm. The proteins myoglobin, myoalbumin, myogen, and droplets of triglyceride and glycogen are present in the sarcoplasm. Myoglobin, or muscle hemoglobin, accepts O_2 from hemoglobin in the blood; thus oxygenated myoglobin serves as a reservoir of O_2 for the muscle fibers. Myogen has the properties of the enzyme aldolase. The function of

myoalbumin is not known. Each striated muscle fiber contains many nuclei and mitochondria, a Golgi apparatus, and a specialized endoplasmic reticulum called the sarcoplasmic reticulum. The sarcoplasmic reticulum is a closed tubular membrane system that extends throughout the sarcoplasm and surrounds the bundles of contractile proteins within each myofibril. A three-dimensional model of the sarcoplasmic membrane systems in the rat diaphragm is shown in Figure 21.3. It is composed of longitudinal sections that run the length of each sarcomere and the terminal cisternae portions, which are observed in electron micrographs as triads at the level of the Z band. The other intracellular membrane in muscle is the transverse tubule system which occurs at the Z band, or in some muscles it lies closer to the A–I junction. The T tubule is a tubular invagination of the sarcolemma which penetrates to the interior of the muscle fiber and is considered to be extracellular space. The terminal cisternae portions of the sarcoplasmic reticulum abut the transverse system;

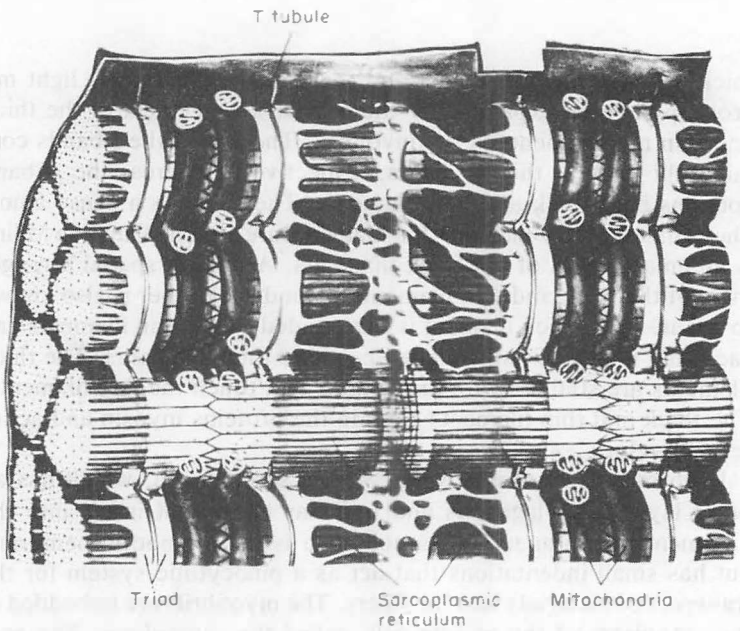


Figure 21.3
Three-dimensional model of the sarcoplasmic membrane system in the rat diaphragm.
Reprinted with permission from H. Schmalbruch, *Advan. Anat., Embryol. Cell. Biol.*, 43(1), 1970.

however, the two membranes remain separated by a section of sarcoplasm.

At least two types of striated muscle fibers are found in skeletal muscle—red fibers and white fibers. Other types of fibers, intermediary between the red and white varieties, also have been detected. Red fibers have a large amount of sarcoplasm, and contain more nuclei and mitochondria than white fibers. Red fibers also contain a greater amount of myoglobin, mitochondrial iron-containing enzymes, and more lipid droplets than white fibers. Red fibers twitch longer, are more easily tetanized, and contract more slowly than white fibers. Three types of skeletal muscle fibers have been designated based on their metabolic and physiological characteristics. The fast-twitch white muscle fibers have a low respiratory capacity, a high glycogenolytic capacity and a high myosin ATPase activity; the fast-twitch red fibers have a high respiratory capacity, a high glycogenolytic capacity, and high myosin ATPase activity. The slow-twitch red fibers have a high respiratory capacity, a low glycogenolytic capacity and low myosin ATPase activity. (See Clin. Corr. 21.1.)

Contractile Proteins

Myosin is the most abundant protein in skeletal muscle, accounting for 60–70% of total protein, and is the major protein of the thick filaments. Myosin is composed of two identical heavy chains about 200,000 mol wt, and four light chains of about 20,000 mol wt. Myosin has a double-headed globular region joined to a double stranded α-helical rod. The schematic representation of the myosin molecule is shown in Figure 21.4. The globular portions of the myosin have

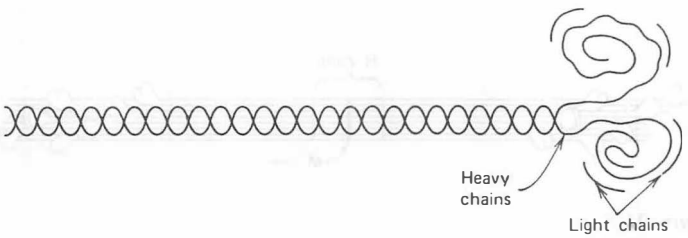


Figure 21.4
Schematic representation of the myosin molecule.
The myosin molecule consists of two heavy chains and four light chains. Each heavy chain has a molecular weight of ~200,000 and has an α-helical portion and a globular head. The four light chains are bound to the two globular heads of the heavy chains.

CLIN. CORR. 21.1 DUCHENNE'S MUSCULAR DYSTROPHY

A number of components in the sarcoplasm of skeletal muscle are found to be present in lower than normal levels in patients with Duchenne's muscular dystrophy: glycolytic enzymes, creatine phosphokinase, adenosine monophosphate deaminase, ATP, phosphocreatine, K^+, and myoglobin. Most of the biochemical abnormalities have been interpreted as a failure to retain sacroplasmic substances within the muscle fiber and *cannot* be stated to be specific inborn errors of metabolism.

ATPase activity and will combine with actin. Two of the light chains (one on each globular head) are identical and may be removed with no loss of ATPase activity. The other two light chains are not identical and are required for both the ATPase activity and the actin binding properties of the myosin. A region of the α-helical portion of the myosin molecule adjacent to the globular portion of the molecule is thought to be flexible. A thick myofilament consists of about 400 rods of myosin bound together as depicted in Figure 21.5. The midpoint at which the myosin fibers meet tail to tail is at the M line.

About 20–25% of total muscle protein is the globular protein, called actin. The globular form is called G-actin and consists of a single peptide chain. F-actin consists of a double-stranded helix of G-actin monomers forming a thin filament. A complex called actomyosin is formed when a solution of actin is mixed with a solution of myosin. Strands of actomyosin will contract in the presence of ATP.

Tropomyosin is a rod-shaped molecule found associated with the actin filaments. Tropomyosin consists of two similar α-helical peptide chains coiled around each other in a head-to-tail assembly and is attached to the chains of F-actin. Troponin is a spherical molecule, found bound to the actin filaments, and consists of three different subunits. The probable relative positions of tropomyosin, troponin, and actin in the thin filament of muscle is shown in Figure 21.6. A calcium binding protein (TN–C) has two high and two low affinity Ca^{2+} binding sites and is the only subunit of troponin to bind Ca^{2+}. The binding of Ca^{2+} to TN–C induces a conformational change in this subunit, which is recognized by both tropomyosin and actin. The inhibitory protein (TN–I) inhibits the interaction of actin with myosin and also inhibits ATPase activity. Tropomyosin binding sub-

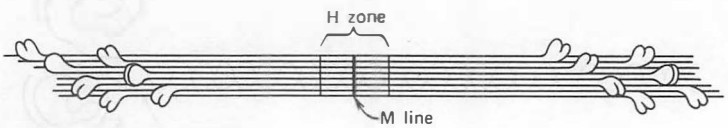


Figure 21.5
Schematic representation of a thick filament formed by the assembly of the individual myosin molecules (Figure 21.4).
The α-helical ends of the myosin heavy chains are joined together in a staggered array, with the result that every fourth pair of myosin heads are aligned with each other. The midpoint of the aggregate of the individual molecules (~400) is the M line, which is in the middle of the H zone.

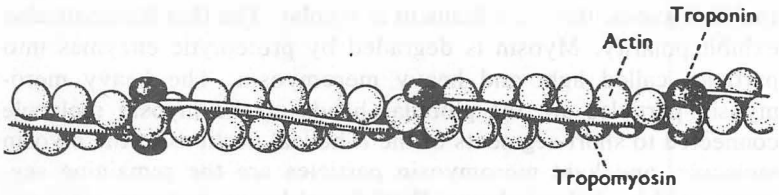


Figure 21.6
Probable relative positions of tropomyosin, troponin, and actin in the thin filament of muscle.
Tropomyosin lies in the groove of the helix, which is undisturbed. There is one tropomyosin and one troponin for each seven pairs of G-actin.
Reprinted with permission from H. E. Huxley, Regulation of muscle function by tropomyosin–troponin, in Y. Hatefi and L. Djavadi-Ohaniance, eds., *The Structural Basis of Membrane Function,* Academic Press, New York, 1976, p. 319.

unit (TN–T) mediates the binding of the TN–I and TN–C subunits to the actin–tropomyosin complex and also the binding of TN–I to TN–C. α-Actinin is associated with the Z line in the sarcomere and is the protein that binds the actin filaments to the Z line. γ-Component is found associated with the troponin molecule and has now been identified as phosphocreatine kinase. C-protein is isolated from preparations of myosin and is involved in the assembly of the myosin molecules into the thick filament. β-actinin is found associated with the F-actin molecules and is a length-determining factor for the assembly of the thin filaments. M-protein is detected in the M line. The approximate molecular weights of skeletal muscle contractile proteins are summarized in Table 21.1.

Table 21.1 Molecular Weights of Skeletal Muscle Contractile Proteins

Myosin	500,000
Heavy chain	200,000
Light chain	20,000
Actin monomer (G-actin)	42,000
Tropomyosin	70,000
Troponin	76,000
TN–C subunit	18,000
TN–I subunit	23,000
TN–T subunit	37,000
α-Actinin	200,000
C-protein	150,000
β-Actinin	60,000
M-protein	100,000

Theory of Muscle Contraction

Skeletal muscle contracts as the length of the sarcomeres decreases, and this is accomplished by the actin filaments sliding past the myosin filaments into the H zone of the sarcomere. The sliding filament model of muscle contraction involves the ATPase activity of the globular portion of the myosin molecule, its ability to bind to actin, and the manner in which the molecules of myosin aggregate to form the thick filament. The fact that the globular heads of the myosin molecules in the filament are pointed in opposite directions on either side of the M line of the sarcomere means there is a directionality of the contractile force. The binding of the heads of myosin with the actin, forming cross-linkages, will exert a force on one side of the A band that is opposite to that on the other side of the A band.

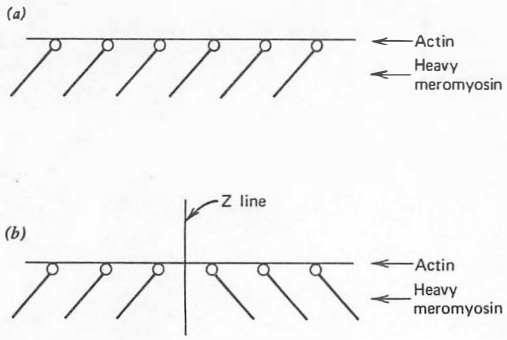


Figure 21.7
Schematic representation of the attachment of heavy meromyosin to actin filaments isolated free (a) and attached to the Z line (b).
Particles of heavy meromyosin combine with actin filaments with the α-helical portions of the particles orientated in the same angular direction from the actin filament (a). Heavy meromyosin particles combine with actin filaments attached to the Z line with the α-helical portions orientated in opposite angular directions on either side of the Z line (b).

In other words, the thick filament is bipolar. The thin filaments also exhibit polarity. Myosin is degraded by proteolytic enzymes into particles called light and heavy meromyosin. The heavy meromyosin particles are the globular heads of the myosin molecule connected to short segments of the α-helical portions of the myosin molecule, and light meromyosin particles are the remaining segments of the α-helical chains. Particles of heavy meromyosin combine with actin filaments with the globular myosin heads attached to the actin filaments, and with the α-helical portions of the heavy meromyosin orientated in the same angular direction from the actin filament. Actin filaments, connected to the Z line, combine with heavy meromyosin with the α-helical portions orientated in opposite angular directions on either side of the Z line. Figure 21.7 is a schematic representation of the attachment of heavy meromyosin to actin filaments isolated free (*a*) and to actin filaments attached to the Z line (*b*). Since the thin filaments on one side of the Z line have the same orientation, whereas those on the opposite side have the reverse polarity, the force generated by the myosin heads binding to the actin is in the direction of the M line on either half of the sarcomere.

The events thought to take place in skeletal muscle contraction are as follows: An electrical impulse from the motor nerve is transmitted to the muscle at the neuromuscular junction. This potential is transferred across the synapse by the release of acetylcholine. The impulse spreads over the entire sarcolemma, which then becomes depolarized. A potential difference of about 60 mV exists between the inside and outside of the resting muscle cell (positive outside). As the impulse spreads over the sarcolemma, the potential disappears as Na^+ enters the sarcoplasm, followed by a loss of K^+ to the exterior of the cell. The transverse tubule also becomes depolarized and transmits the impulse to all of the myofibrils within the fiber. The depolarization impulse is transmitted to the adjacent terminal cisternal portions of the sarcoplasmic reticulum from the transverse tubules in some unknown manner, and the longitudinal sarcoplasmic reticulum rapidly discharges Ca^{2+} into the sarcoplasm. The Ca^{2+} concentration of the sarcoplasm is increased by a factor of at least 10-fold. The Ca^{2+} binds to the TN–C subunit of troponin, which undergoes a conformational change that is recognized by tropomyosin. The tropomyosin moves, relative to the actin, which is thought to expose the myosin binding sites on the actin. The actomyosin Ca^{2+},Mg^{2+}-ATPase becomes activated and the energy produced by hydrolysis of ATP is used for the attachment and detachment of actin and myosin to force the actin filaments into the H zone. The mechanism of the force-generating step is not known, but probably involves conformational changes in the flexible region of the

myosin molecule near its globular head. Relaxation of the skeletal muscle occurs with cessation of the nervous impulse and the return of the sarcolemma and transverse tubule to its original polarized state. That is, Na^+ and K^+ are pumped out and into the cell, respectively, by the Na^+-K^+ ATPase, or called the "sodium pump." The permeability of the sarcoplasmic reticulum to Ca^{2+} decreases, and Ca^{2+} is rapidly transported from the sarcoplasm into the sarcoplasmic reticulum by the ATP-dependent Ca^{2+} pump.

Role of the Sarcoplasmic Reticulum in Muscle Contraction and Relaxation

The importance of the sarcoplasmic reticulum in furnishing Ca^{2+} for muscle contraction is well established. Sarcoplasmic reticulum from skeletal muscle is readily isolated and will reseal into closed vesicles that can be isolated by differential centrifugation. The vesicles retain the ability to pump Ca^{2+} which is coupled to ATP hydrolysis by a membrane-bound Ca^{2+} and Mg^{2+}-dependent ATPase.

Four proteins are found in skeletal muscle sarcoplasmic reticulum. The major protein is a Ca^{2+},Mg^{2+}-dependent ATPase, which has two high affinity Ca^{2+} binding sites. The ATPase forms a phosphoprotein intermediate when hydrolyzing ATP, and its decomposition is activated by Mg^{2+} as follows:

$$E + ATP \xrightarrow{Ca^{2+}} E{-}P + ADP$$

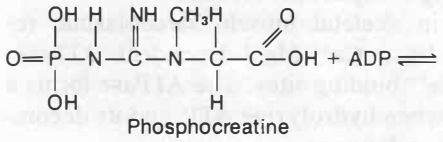
$$E{-}P + H_2O \xrightarrow{Mg^{2+}} E + P_i \text{ (inorganic phosphate)}$$

A low molecular weight (~12,000) proteolipid, rich in arginine and glutamic acid and covalently bonded to two fatty acids, may be involved in mediating the transport of Ca^{2+} across the lipid bilayer. Two calcium binding proteins have been isolated that have no enzymatic activity. An acidic protein called calsequestrin has one-half of the amino acid residues as glutamic acid and aspartic acid. Another Ca^{2+} binding protein is less acidic then the calsequestrin and binds about one-half as much Ca^{2+} as calsequestrin.

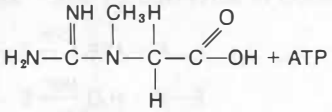
A tentative model for the Ca^{2+} transport system in sarcoplasmic reticulum involves the ATPase and proteolipid as intrinsic membrane proteins. The ATPase has both polar and nonpolar domains and the nonpolar portion (possibly containing a Ca^{2+} ionophore site) is in the lipid bilayer. The polar portion is on the surface of the membrane and is the site for the hydrolysis of ATP. The interaction between the site of ATP hydrolysis and the portion of the ATPase buried in the lipid bilayer probably controls the transport of Ca^{2+} across the membrane. The Ca^{2+} binding proteins are extrinsic proteins bound to the membrane inside the lumen of the sarcoplasmic reticulum and act to store the Ca^{2+} in the sarcoplasmic reticulum.

Energy Source for Muscle Contraction

The energy for muscle contraction is derived from the hydrolysis of ATP. The concentration of ATP in resting muscle will supply sufficient energy for contraction for only a fraction of a second, and the energy demands of intense contraction will exceed the capacity of the muscle to generate sufficient ATP by the metabolism of its various metabolic fuels. Mammalian muscle contains a reserve store of "high energy phosphate" in the form of phosphocreatine. Resting muscle contains six times as much phosphocreatine as ATP, and whenever the expenditure of ATP exceeds its production, the phosphocreatine stores are used in an attempt to replenish the ATP supplies. The production of ATP from metabolic fuels is greatly increased in muscle as the work load is increased. The phosphocreatine stores are therefore considered a "backup system" for the generation of ATP. The phosphorylation of ADP by phosphocreatine is catalyzed by the enzyme phosphocreatine kinase as follows:



Phosphocreatine



Creatine

After the completion of a contraction, or when the generation of ATP (by oxidation of metabolic fuel) in the muscle exceeds its utilization, phosphocreatine kinase catalyzes the phosphorylation of creatine by ATP. A second "backup system" for the generation of additional ATP is the enzyme myokinase that catalyzes the following reaction:

$$2ADP \longrightarrow ATP + AMP$$

Metabolic Fuel of Skeletal Muscle

After a Meal

The metabolic fuel of skeletal muscle after a meal is glucose, at which time the glucose and insulin levels in the blood are high, and free fatty acid levels are low. Glucose is metabolized principally by the glycolytic pathway, with less than 2% of glucose metabolized by the hexose monophosphate shunt. The glycolytic pathway and the citric acid cycle operate at ~10 and 5–10% of their maximal capacities, respectively, in resting muscle. Most of the products of

glycolysis are completely oxidized in noncontracting muscle with only small amounts of lactate entering the blood. The rates of glucose uptake and lactate production by the skeletal muscle are stimulated in anoxia. The stimulation of the glycolytic rate in anoxia is considered to be via the activation of phosphofructokinase, pyruvate kinase, and glyceraldehyde 3-phosphate dehydrogenase. Both ATP and phosphocreatine inhibit all three enzymes and the amounts of both compounds are low in anoxic muscle. The enzyme phosphofructokinase is also activated by a variety of compounds, for example, AMP, ADP, inorganic phosphate, and NH_4^+, that are found in increased amounts in skeletal muscle after anoxia. During contraction of skeletal muscle in the presence of excess glucose, the consumption of both glucose and oxygen increases about 20-fold. The stimulation of the glycolytic rate is by the same mechanisms that occur in anoxia. The citric acid cycle activity is stimulated through the increased activity of isocitrate dehydrogenase, which is inhibited by ATP and activated by AMP. As ATP and AMP levels decrease and increase, respectively, the flux of metabolites through the citric acid cycle increases. In prolonged heavy contraction, the rate of the production of pyruvate by glycolysis exceeds its rate of metabolism by the citric acid cycle, and large amounts of lactate are transported to the liver. The fate of the lactate remaining in muscle after contraction has ceased has long been a matter of controversy, but it is now believed that fast-twitch red and white types of skeletal muscle can synthesize glycogen from lactate. In contrast, the slow-twitch red type of muscle has a limited capacity of glycogen synthesis because of a lack of fructose 1,6-bisphosphatase.

Insulin stimulates the glucose uptake and oxygen consumption of muscle and also stimulates the uptake of amino acids and the accumulation of protein. The stimulation of oxygen consumption is by an activation of pyruvic dehydrogenase, whereas the insulin effect on incorporation of amino acids is presumably indirect, in that insulin inhibits the breakdown of muscle protein.

Large amounts of NADH are produced in the sarcoplasm when glucose is the metabolic fuel. Some NADH is reoxidized in the reduction of pyruvate to lactate, especially in heavy contraction, and the remainder by the transfer of the reducing equivalents to the mitochondria via a substrate shuttle pathway (see page 296).

In Fasting

The metabolic fuel of skeletal muscle in the early stages of fasting are free fatty acids mobilized from adipose tissue which are the preferred fuel, since the presence of the free fatty acids actually supresses the uptake and oxidation of glucose. The free fatty acids promote indirectly the conversion of pyruvic dehydrogenase to its

inactive form, thus inhibiting glucose utilization. Fatty acids are estimated to supply at least 50–60% of the energy needs of skeletal muscle both at rest and in contraction, and, unlike glucose, the utilization of fatty acids by muscle does not require insulin. The metabolic fuel of skeletal muscle in later stages of fasting are the ketone bodies, acetoacetate and β-hydroxybutyrate, synthesized by the liver. Ketone bodies, like fatty acids, promote the conversion of pyruvic dehydrogenase to its inactive form. The utilization of fatty acids and ketone bodies by muscle spares glucose for tissues that utilize glucose as their principle metabolic fuel. The removal of ketone bodies by the muscle helps reduce acidosis or ketosis, and, of course, salvages energy that would be lost if the ketone bodies were excreted in the urine. Muscle cannot utilize fatty acids or ketone bodies for energy in anoxia or ischemia.

The branched-chain amino acids (leucine, isoleucine, and valine) are degraded in fasting to yield energy in extrahepatic tissues, notably skeletal and cardiac muscle. Both fatty acids and epinephrine stimulate the oxidation of the branched-chain amino acids by skeletal muscle. Thus the action of epinephrine on skeletal muscle branched-chain amino acid utilization may well be indirect via its lipolytic action on adipose tissue. The first step in the catabolism of the branched-chain amino acids is transamination with pyruvate to form alanine, which is released from muscle in fasting to be converted to glucose in liver and kidney (page 607).

In Tetanic Contraction

The total amount of glycogen in skeletal muscle far exceeds the amount of glycogen in liver. The amount of glycogen per unit weight in muscle is ~10% of the amount found in liver. However, the mass of skeletal muscle is over 25 times greater than liver. The glycogen in muscle is used as its energy source in anoxia and tetanic contraction. Muscle glycogen, in contrast to liver glycogen, is not depleted in fasting and is not broken down to glucose for use by other tissues because muscle lacks the enzyme glucose 6-phosphatase. At least two different systems have been found for the activation of the phosphorylase in muscle. Epinephrine activates the phosphorylase system in muscle (see page 382), but the time required for this mechanism is about 3 min, which is relatively slow in comparison to the activation of the phosphorylase induced by contraction. Contraction-induced phosphorylase activation is by the stimulatory action of Ca^{2+} on both activated and nonactivated phosphorylase kinase and on the autoactivation of phosphorylase kinase. The hormonal and contraction-induced activation of skeletal muscle phosphorylase is shown in Figure 21.8.

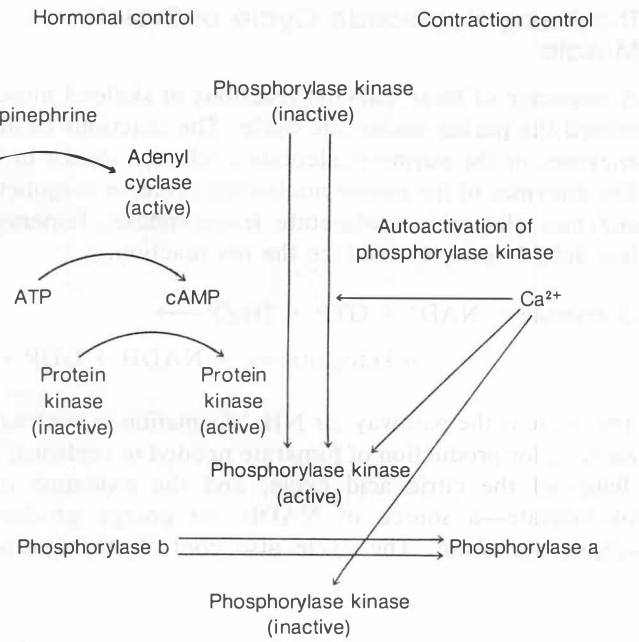


Figure 21.8
The hormonal- and contraction-induced activation of skeletal muscle phosphorylase.
The contraction-induced activation of phosphorylase is by the stimulatory effect of Ca^{2+} on the autoactivation of phosphorylase kinase and on the action of both active and inactive phosphorylase kinase.

Both the I and D forms of glycogen synthase are found in muscle, and only the I form (dephosphorylated) is considered to be active in vivo. The enzyme that activates the glycogen synthase is a phosphoprotein phosphatase, and it is now thought that this phosphatase also inactivates glycogen phosphorylase by a dephosphorylation. The inactivation of glycogen synthase by a phosphorylation is a complex reaction involving multiple phosphorylation sites on the synthase and utilizing two or more protein kinases (perhaps four). One protein kinase is a cAMP-dependent protein kinase, which is activated by glucagon. A cAMP-*independent* protein kinase also is found in skeletal muscle, which is activated by a calcium-dependent regulator protein. The Ca^{2+}-dependent regulator protein has no effect on the cAMP-dependent protein kinase.

The Purine Nucleotide Cycle of Skeletal Muscle

A sequence of three enzyme reactions in skeletal muscle has been named the purine nucleotide cycle. The reactions catalyzed by the enzymes in the purine nucleotide cycle are shown in Figure 21.9. The enzymes of the purine nucleotide cycle, in conjunction with the enzymes glutamate-oxalacetate transaminase, fumarase, and malate dehydrogenase, catalyze the net reaction,

$$\text{Glutamate} + NAD^+ + GTP + 2H_2O \longrightarrow$$
$$\alpha\text{-ketoglutarate} + NADH + GDP + NH_4^+ + P_i$$

The cycle is the pathway for NH_4^+ formation in contracting skeletal muscle, for production of fumarate needed to replenish the intermediates of the citric acid cycle, and the oxidation of malate to oxalacetate—a source of NADH for energy production by the respiratory chain. The cycle also controls the levels of adenine

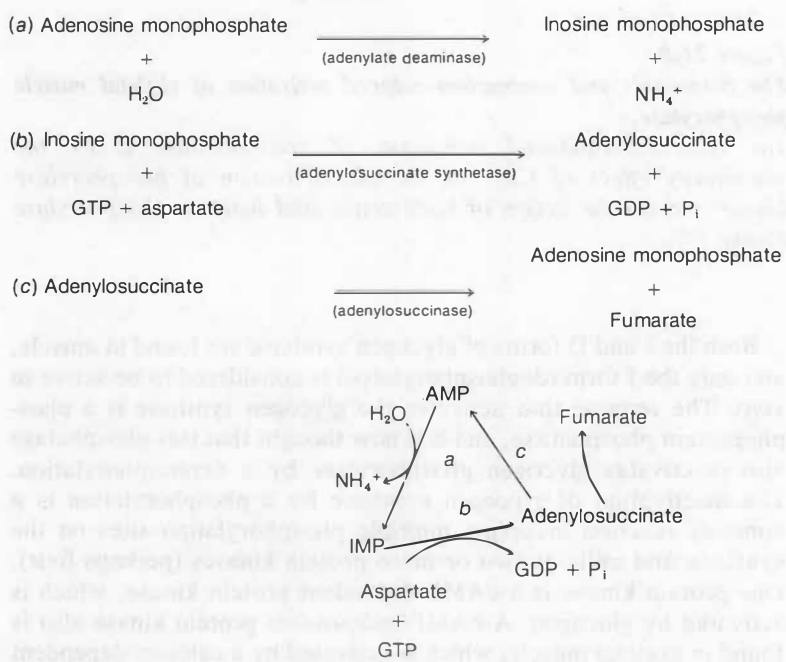


Figure 21.9
Reactions catalyzed by enzymes in the purine nucleotide cycle.

nucleotides, inorganic phosphate, and NH_4^+, all of which have an effect on the rate of glycolysis at the level of phosphofructokinase.

Effect of Training Exercise on Skeletal Muscle Metabolism

Both the size and number of skeletal muscle mitochondria are greater in exercise-trained muscle than in nontrained muscle. Trained skeletal muscle has greater amounts of the enzymes used in fatty acid oxidation and ketone body utilization than does nontrained muscle and therefore has a greater capacity to utilize noncarbohydrates for energy. Thus the onset of the depletion of body carbohydrate, accumulation of lactate, and exhaustion during prolonged muscle contraction is delayed longer in trained muscle than in nontrained muscle.

21.2 CARDIAC MUSCLE

Cardiac muscle is striated and resembles red skeletal muscle but the fibers are more irregular in shape and have a variegated appearance caused by the branching of the fibers. Despite the branching, cardiac muscle fibers do not form an irregular network but are arranged in layers within which they tend to be parallel. The contractile proteins described for skeletal muscle are present in cardiac muscle and perform the same functions, even though the composition and the structures of the contractile proteins from the two tissues are not identical.

Regulation of Cardiac Muscle Contraction

Both the force and the rate of the contraction of the heart are stimulated by epinephrine, which stimulates the production of cAMP and leads to an increase in a protein kinase activity. A number of heart proteins are phosphorylated by the protein kinase, and all are thought to be involved in the control of cardiac contraction. The events in modulating cardiac contraction are suggested to be as follows. Epinephrine stimulates the formation of cAMP, which stimulates the cAMP-dependent protein kinase. The protein kinase catalyzes the phosphorylation of the myofibrillar proteins TN–I and myosin light chain, a sarcoplasmic membrane protein, phospholamban, and sarcolemma membrane proteins. Calcium concentration in the sarcoplasm is increased by entry of Ca^{2+} through the sarcolemma and Ca^{2+} release from the sarcoplasmic reticulum. After

the contraction, the Ca^{2+} is sequestered back into the sarcoplasmic reticulum to be released at the time of the next contraction.

Metabolic Fuel of Cardiac Muscle

A variety of metabolic fuels can be utilized by the heart, but 60–90% of the total oxidative metabolism is accounted for by oxidation of fatty acids. Glucose, pyruvate, and lactate are utilized preferentially after a meal when the levels of free fatty acids in the blood are low. Both pyruvate and lactate inhibit the uptake and oxidation of free fatty acids and of the branched-chain amino acids, isoleucine, leucine, and valine. Fatty acids mobilized from the adipose tissue are the major metabolic fuel of heart in fasting; a lipoprotein lipase of heart is stimulated in fasting, increasing the hydrolysis of plasma triglycerides. In addition, branched-chain amino acids are utilized for energy in short-term fasting. In long-term fasting, the ketone bodies become the preferred fuel for the heart. The utilization of glucose and endogenous glycogen by the heart increases about 10- to 20-fold in anoxia because the fatty acids and ketone bodies cannot be utilized for energy. Acetate, present in blood in significant amounts only after the ingestion of alcohol, is also used as an energy source for heart.

The inhibition of the utilization of glucose and pyruvate by fatty acids and the ketone bodies involves a number of control mechanisms. Fatty acids inhibit glucose entry into the cell and also inhibit the enzyme phosphofructokinase. The accumulation of glucose 6-phosphate, due to the inhibition of phosphofructokinase, leads to an inhibition of hexokinase. The utilization of the fatty acids and ketone bodies by heart also elevates the levels of NADH and acetyl CoA in the mitochondria. Mitochondrial isocitrate dehydrogenase and α-ketoglutarate dehydrogenase are inhibited by the unavailability of their coenzyme, NAD^+, and as a result of their inhibition, the levels of citrate in the mitochondria become elevated. The excess citrate enters the sarcoplasm, where it inhibits phosphofructokinase. The high levels of acetyl CoA and NADH in the mitochondria inhibit pyruvate dehydrogenase by a feedback type of enzyme inhibition. NADH and acetyl CoA also stimulate a kinase, which catalyzes the phosphorylation of pyruvate kinase to an inactive form.

The rate-controlling step in the utilization of the branched-chain amino acids by heart is the α-keto acid dehydrogenase, which utilizes as its substrates the α-keto acids from leucine, isoleucine, and valine. The inhibition of branched-chain amino acid utilization by pyruvate in the heart is attributed to a pyruvate inhibition of the α-keto acid dehydrogenase. The mechanism of the inhibition of fatty acid oxidation by pyruvate is not known.

Table 21.2 The Metabolic Fuel of Heart in Various Physiological States

Physiological State	*Metabolic Fuel*
After a meal	Glucose, lactate, pyruvate
Fasting (short-term)	Free fatty acids, triglycerides, leucine, isoleucine, valine
Fasting (long-term)	Acetoacetate, β-hydroxybutyrate
Anoxia	Glucose, glycogen
Heavy work load	Free fatty acids
Alcohol ingestion	Acetate

The stimulation of glucose utilization in anoxia is by the mechanisms described in skeletal muscle. Three mechanisms have been suggested for the stimulation of the breakdown of heart glycogen in anoxia: an epinephrine activation of the phosphorylase system, an inhibition of phosphorylase phosphatase by the large amounts of AMP formed in anoxia, and by an activation of phosphorylase b by AMP. The supplies of cardiac glycogen are replenished after the anoxia by an activation of the glycogen synthase, which is activated (dephosphorylated) by a specific phosphatase.

An increase in the level of mechanical work by the heart is accompanied by an increase in the oxygen consumption, with a concomitant increase in oxidative phosphorylation and a shift from glucose to palmitate utilization. The carnitine acetyltransferase system plays a role in the coupling of the rate of fatty acid uptake by cardiac mitochondria to the activity of the citric acid cycle. In other words, the rate of translocation of acyl units across the inner mitochondrial membrane limits the rate of long-chain fatty acylcarnitine oxidation.

A summary of the metabolic fuel of heart in various physiological states is shown in Table 21.2.

21.3 ADIPOSE TISSUE

The Role of Adipose Tissue in Mammals

"Body fat" has been considered for many years to be connective tissue filled at random with droplets of fat, but when examined by electron microscopy, it becomes apparent that it is a specialized

mammalian tissue. Adipose tissue has a central role in the energy metabolism of the entire animal. The end products of the digestion of dietary fat, carbohydrate, and protein are converted to triglycerides by the adipose tissue after a meal. The adipose triglycerides are hydrolyzed to glycerol and fatty acids when the animal is in the fasting state, and the fatty acids are used by a variety of other tissues as their metabolic fuel. The triglycerides of adipose tissue are the major store of metabolic fuel for the body and are also a major source of energy for heat production. The oxidation of fatty acids yields more than twice the energy of carbohydrate or protein per unit weight. The deposition of the triglyceride in specialized cells eliminates the need for extensive storage of carbohydrate or triglyceride in the other tissues that might interfere with their function. The amount of adipose triglyceride in the normal average adult human is about 25 lb (~11 kg), sufficient to maintain life for 40 days.

Distribution and Chemical Composition of Adipose Tissue

Adipose tissue is widely distributed in the body, in muscles, under the skin, around blood vessels, and in the abdominal cavity. White adipose tissue is involved in energy storage, while brown adipose tissue is involved in heat production of the animal. White adipose tissue cells are spherical, containing a large vacuole of triglyceride that occupies almost the entire cell, and the cytoplasm surrounding the lipid vacuole is no more than a thin film, invisible by light microscopy. A small number of mitochondria are found in the cytoplasm along the periphery of the central lipid droplet, a flattened Golgi apparatus is near the nucleus, and an endoplasmic reticulum is present.

Triglyceride constitutes ~80% of the wet weight of human white adipose tissue and ~99% of the lipid. Approximately 20 fatty acids are found in human adipose triglyceride, the principal ones being oleic acid (45%), palmitic acid (20%), linoleic acid (10%), stearic acid (6%), and myristic acid (4%).

Brown adipose tissue cells are polygonal, and the cytoplasm is more abundant and granular than that in white adipose cells. The cells contain a number of small lipid droplets, and the nucleus is not flattened as it is in the white adipose cells. The cytoplasm contains small amounts of endoplasmic reticulum and Golgi apparatus and numerous mitochondria. Both white and brown adipose tissue are well supplied with blood capillaries and are well innervated.

Adipose Tissue Metabolism After Feeding

Triglycerides in the chylomicrons and in the very low density lipoproteins (VLDL) are hydrolyzed to fatty acids and glycerol by lipo-

protein lipase, which is secreted from the adipocyte and becomes associated with blood capillary walls, where it exerts its catalytic effect on the triglycerides. Insulin activates lipoprotein lipase in some unknown manner, thus the enzyme is considered a control step in the assimilation of fatty acids from triglycerides into adipose tissue. The fatty acids enter the adipocyte, where they are esterified to form triglycerides, and the glycerol is transported to the liver, where it is used for the synthesis of glucose. The lipoprotein particles remaining after the action of lipoprotein lipase on VLDL are called remnants and are transformed to low density lipoproteins (LDL) and high density lipoproteins (HDL) in the liver.

Most of the glucose taken up by adipose tissue after a meal is used for the synthesis of triglycerides, with small amounts converted to glycogen. This is in contrast to the fate of glucose in liver, where most of the glucose is used for the synthesis of glycogen. The glycogen that is formed in adipose tissue after feeding is later converted to triglyceride. The control of glycogen synthesis in adipose tissue is via the interconversion of the D form of glycogen synthase (phosphorylated) to the I form (nonphosphorylated).

Glycogen synthesis is stimulated after feeding and insulin is involved in some manner. The rate of entry of glucose into the adipocyte is a rate-determining step in glucose metabolism and is in some way stimulated by insulin. The uptake of glucose by adipocytes is by a process of facilitated diffusion (passive mediated transport), and two glycoproteins have been isolated from adipocyte plasma membranes that may be the membrane carrier. Various proteolytic enzymes have an insulin like activity on glucose transport in the adipocyte, and a proteolytic activity of insulin has been reported, suggesting that insulin might increase glucose transport involving a proteolytic process. Several compounds that have sulfhydryl groups inhibit adipocyte glucose transport activity, and the treatment of adipocytes with compounds that react with sulfhydryl groups abolishes the ability of insulin to stimulate adipocyte glucose transport. Insulin has been reported to inhibit the phosphorylation of adipocyte membrane proteins by ATP; thus an insulin interaction with its plasma membrane receptor has been suggested to generate a second messenger, which regulates the phosphorylation of membrane proteins. A number of oxidants, including H_2O_2, stimulate glucose transport in the adipocyte, and the stimulation of H_2O_2 production in adipocytes by insulin has been found. The insulin effect on glucose transport may be, in part, an indirect effect mediated by its stimulation of intracellular rates of glucose utilization. The mechanism of the stimulation of adipocyte glucose transport by insulin is the subject of intensive investigation in a number of laboratories.

The phosphorylation of glucose to glucose 6-phosphate by adipose hexokinase is under the control of insulin and or glucose. Adipose

tissue from fed rats has higher hexokinase activities than adipose tissue from fasted rats. Approximately 23% of glucose taken up by the adipocyte is metabolized by the hexose monophosphate shunt, and the remainder is metabolized by glycolysis. The insulin-induced stimulation of both glucose transport into the adipocyte and of hexokinase activities produces an increase in the glycolytic rate after feeding. The reasons for the increase in glucose metabolism by the hexose monophosphate shunt after feeding are not known. The increase in the glycolytic rate and in hexose monophosphate shunt metabolism after feeding increases the production of NADH and NADPH, respectively. Some of the NADH is used for the reduction of dihydroxyacetone phosphate to glycerol 3-phosphate, and the remaining NADH is transhydrogenated to form NADPH ($NADH + NADP^+ \rightarrow NAD^+ + NADPH$), a reaction that is necessary in adipose tissue because the hexose monophosphate pathway produces only 65% of the NADPH needed for fatty acid synthesis.

The fate of pyruvate in adipose tissue is shown in schematic form in Figure 21.10, and the details of fatty acid synthesis are in Chapter 9. Most of the citrate formed from oxalacetate and acetyl CoA leaves the mitochondria for conversion to acetyl CoA and oxalacetate because the mitochondria have low levels of aconitase and isocitrate dehydrogenase activities that restrict the metabolism of citrate by the citric acid cycle. The transfer of citrate out of the mitochondria and the conversion of oxalacetate to glycerol 3-phosphate means there is a loss of oxalacetate from the mitochondria. The mitochondria are impermeable to cytosolic oxalacetate and the major mechanism for the replenishment of mitochondrial oxalacetate is the carboxylation of pyruvate, which comes from the reduction of oxalacetate to malate and the decarboxylation of malate via malate dehydrogenase in the cytoplasm (Figure 21.11).

The increase in the rate of pyruvate formation by glycolysis and in the rate of metabolism of glucose by the hexose monophosphate shunt after feeding are accompanied by increases in the activities of pyruvate dehydrogenase, acetyl CoA carboxylase, citrate cleavage enzyme, and fatty acid synthetase. The increase in pyruvate dehydrogenase activity is the result of its conversion to a dephosphorylated form. The activation of acetyl CoA carboxylase is by the citrate-induced conversion of the monomeric form of the enzyme to its polymeric form, and also by dephosphorylation. The mechanisms of the increase in the activities of citrate cleavage enzyme and fatty acid synthetase are not known. The response of the key control enzymes in fatty acid synthesis in adipose tissue after feeding are listed in Table 21.3.

Table 21.3 The Enzyme Activities in Adipose Tissue That Are Increased in the Fed State as Compared to the Fasting State

Hexokinase
Glucose 6-phosphate dehydrogenase
Acetyl CoA carboxylase
Pyruvate dehydrogenase
Citrate cleavage enzyme
Fatty acid synthetase
Glycerol 3-phosphate acyltransferase

The stoichiometry of the conversion of glucose to palmitic acid

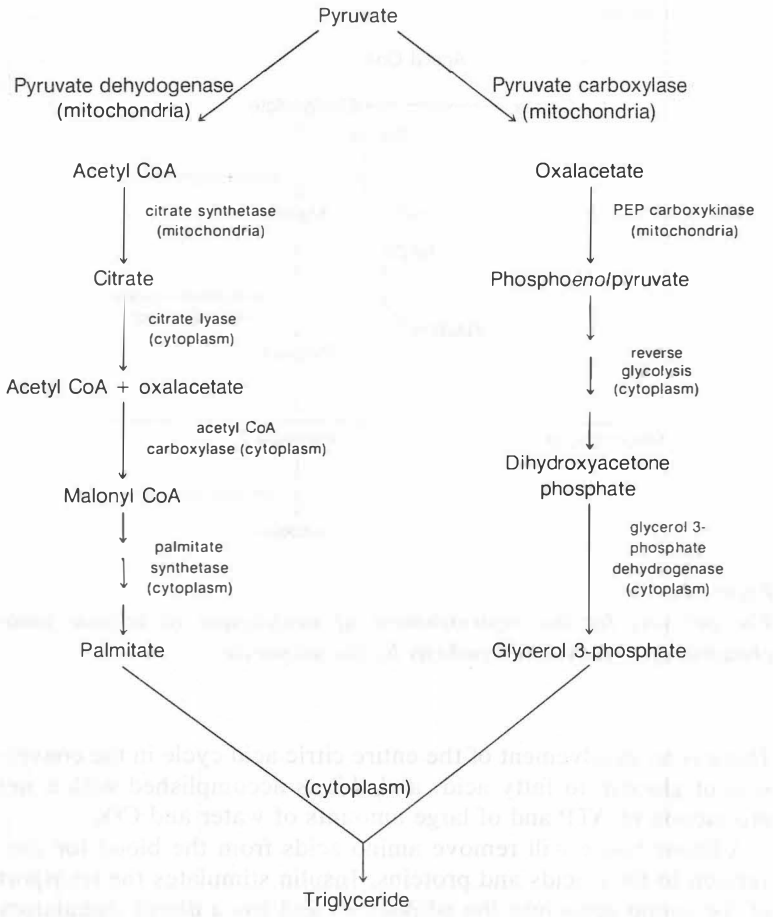


Figure 21.10
The fate of pyruvate in adipose tissue.
The total synthesis of triglyceride from pyruvate occurs as the result of the pathways described in this diagram.

has been calculated—taking into consideration all of the pathways involved: glycolysis; hexose monophosphate shunt; transhydrogenation; conversion of pyruvate to oxalacetate, acetyl CoA, and citrate; cleavage of citrate to oxalacetate and acetyl CoA; and synthesis of palmitate from acetyl CoA. The balanced overall reaction is

$$[4.5\ \text{glucose} + 4O_2 + 9(ADP + P_i) + 8H^+ \longrightarrow$$
$$\text{palmitate} + 11CO_2 + 9ATP + 20H_2O]$$

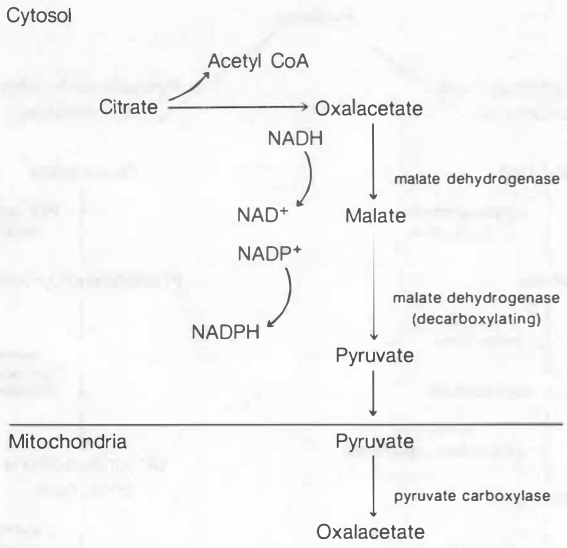


Figure 21.11
The pathway for the replenishment of oxalacetate in adipose mitochondria after fatty acid synthesis by the adipocyte.

There is no involvement of the entire citric acid cycle in the conversion of glucose to fatty acid, and this is accomplished with a net production of ATP and of large amounts of water and CO_2.

Adipose tissue will remove amino acids from the blood for conversion to fatty acids and proteins. Insulin stimulates the transport of the amino acids into the adipocytes and has a direct stimulatory effect on adipocyte protein synthesis. The amino acids differ in their ability to form fatty acids; that is, leucine is, by far, the most potent precursor of all the amino acids for fatty acid synthesis, presumably because the end products of leucine catabolism are acetoacetic acid and acetyl CoA. Acetoacetate also is incorporated into fatty acids in adipose tissue. Adipose tissue converts leucine to cholesterol at a significant rate and this synthesis is stimulated by insulin.

Triglyceride Synthesis in Adipose Tissue

The fatty acids derived from dietary triglycerides or synthesized from glucose and amino acids are used for triglyceride synthesis in the adipocyte. Three pathways of triglyceride synthesis in adipose tissue are (1) glycerol 3-phosphate pathway; (2) monoglyceride pathway; and (3) dihydroxyacetone phosphate pathway. The path-

1. Glycerol 3-phosphate pathway
   Glycerol 3-phosphate + 2 fatty acyl CoA → phosphatidic acid + 2CoA
   Phosphatidic acid → α,β-diglyceride + P_i
   α,β-Diglyceride + fatty acyl CoA → triglyceride
2. Dihydroxyacetone phosphate pathway
   Dihydroxyacetone phosphate + fatty acyl CoA → acyldihydroxyacetone phosphate + CoA
   Acyldihydroxyacetone phosphate → lysophosphatidic acid
   + NADPH + $NADP^+$
   Lysophosphatidic acid + fatty acyl CoA → phosphatidic acid + CoA
   Phosphatidic acid → α,β-diglyceride + P_i
   α,β-Diglyceride + fatty acyl CoA → triglyceride
3. Monoglyceride pathway
   α- or β-Monoglyceride + fatty acyl CoA → α,β-diglyceride
   α,β-Diglyceride + fatty acyl CoA → triglyceride

Figure 21.12
Pathways of triglyceride synthesis in adipose tissue.

ways of triglyceride synthesis are shown in Figure 21.12, with the major route being the glycerol 3-phosphate pathway. The activity of the enzyme catalyzing the acylation of glycerol 3-phosphate (acyl CoA: sn-glycerol 3-phosphate acyltransferase) is increased after feeding. There is a tendency for saturated fatty acids to occupy position 1 and unsaturated fatty acids to occupy position 2 of the triglyceride. A more random pattern of fatty acid distribution is found in position 3 of the triglyceride, with some preference for long-chain fatty acids. The acyltransferase has specificity with respect to the acyl CoA used in the esterification, and a protein "specifier factor" has been found that interacts with acyltransferase(s) to direct the acylation of palmitate to position 2 of the glycerol moiety.

Adipose Tissue Metabolism in Fasting

The triglycerides of adipose tissue are hydrolyzed in the fasting state to glycerol and free fatty acids, by a "hormone-sensitive lipase," which is activated by at least seven hormones (glucagon, epinephrine, ACTH, growth hormone, thyroxine, secretin, and the glucocorticoids). Glucagon and epinephrine are the principal lipolytic hormones in short-term fasting, whereas the glucocorticoids are the lipolytic hormones after long-term fasting or starvation. The "hormone-sensitive lipase" may be a misnomer because this enzyme preparation has at least three other substrates, including cholesterol esters. All of the above listed hormones, except the glucocorticoids, activate an adenylate cyclase in the adipocyte

membrane, and the cAMP activates a protein kinase, which phosphorylates the lipase to its active phosphorylated form. The mechanism of the activation of adipocyte adenylate cyclase by the hormones is complex, and uncertain. The action of epinephrine on adipose hormone-sensitive lipase also is by a process that does not involve cAMP. The mechanism of the activation of adipose triglyceride hydrolysis by the glucocorticoids is not known, but does not involve cAMP. Insulin and/or glucose inhibit fatty acid mobilization from adipose tissue, therefore the mobilization of fatty acids from adipose tissue occurs mainly, if not entirely, in the fasting state. The mechanism of the inhibition of adipose triglyceride hydrolysis by insulin is not certain and may be by an activation of phosphodiesterase which would increase the rate of destruction of cAMP. The prostaglandins also have an antilipolytic effect, in that they reduce the stimulation of lipolysis by epinephrine, glucagon, and so on, and inhibit the induction of cAMP formation by the lipolytic hormones.

The Role of Adipose Tissue in Heat Production

Heat production by brown adipose tissue occurs primarily in newborn animals (including humans) and in adult hibernating animals, while white adipose tissue is involved in heat production in adults, especially during exposure to low temperatures. The production of heat by adipose tissue is instigated by epinephrine, which activates the hydrolysis of adipose tissue triglycerides to fatty acids and glycerol. The released fatty acids have a dual role in heat production, in that they are the source of reducing equivalents (by β-oxidation) for energy production by the respiratory chain of enzymes *and* also act as uncouplers of oxidative phosphorylation. That is, in the presence of the fatty acids, the energy produced by electron transport in the respiratory chain of enzymes is released as heat instead of being converted to ATP. It seems possible that the uncoupling of oxidative phosphorylation could well be accomplished by the fatty acids acting as carrier molecules for protons (proton ionophores) into the mitochondria, thus discharging the proton gradient established by the respiratory chain and bypassing the ATPase involved in the generation of ATP.

21.4 LIVER

The liver has a dual afferent blood supply: the portal vein, carrying blood which has passed through the capillary beds of the alimentary

tract, spleen, and pancreas; and the hepatic artery. A schematic representation of a liver lobule is shown in Figure 21.13. Blood from the branches of these two vessels mixes in passing through the sinusoidal capillaries of the liver lobules. The sinusoids drain into the central veins of the lobules, which are branches of the hepatic veins. The structural unit of liver is the lobule, consisting of mainly parenchymal cells. Branches of the afferent blood vessels together with the bile ducts (portal triad) run along the edges of each lobule. Sinusoidal capillaries pass through the parenchyma from the periphery of the lobule to the central vein. Minute bile capillaries run

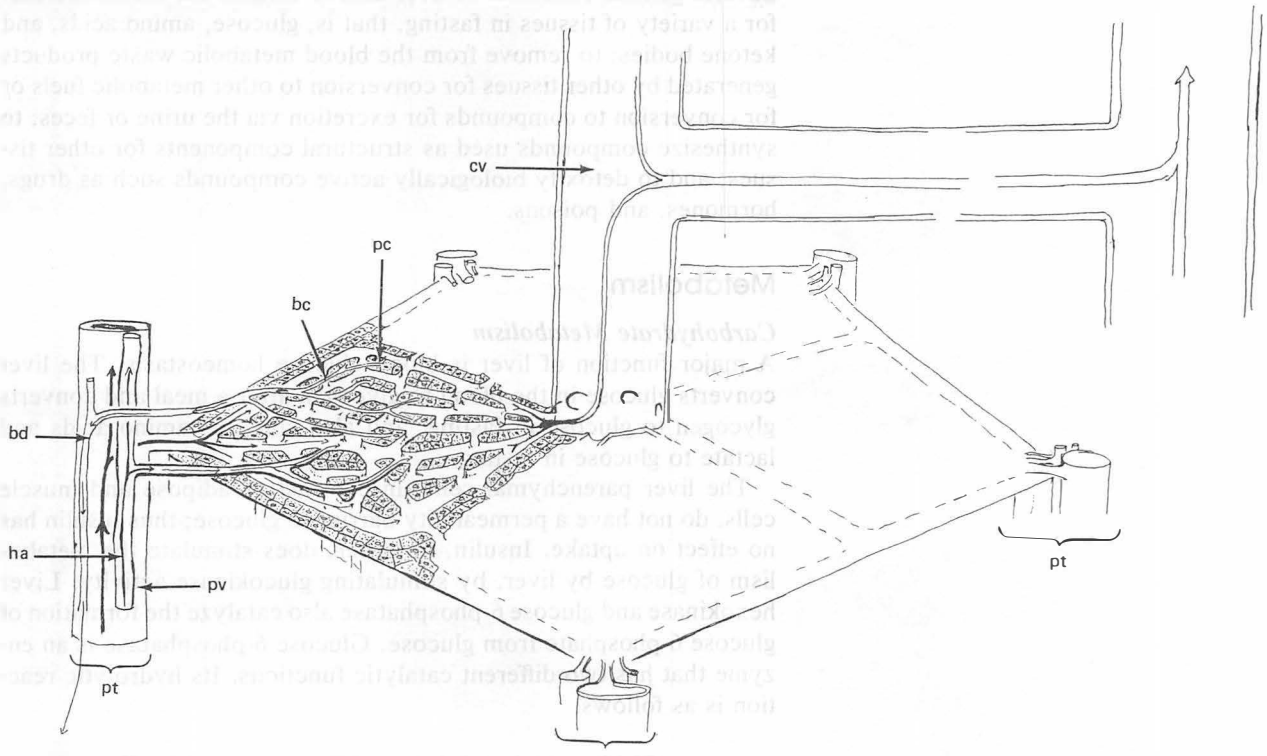


Figure 21.13
A schematic representation of a liver lobule.
Pt is the portal triad, consisting of the portal vein (pv), hepatic artery (ha), and the bile duct (bd). The central vein (cv) drains the lobule, the blood ultimately returning to the vena cava. The bile duct capillaries (bc) leave the parenchymal cells (pc), anastamose at the periphery of the lobule and enter the bile ducts.
Drawing by Dr. Donald W. Robertson, Department of Anatomy, College of Medicine, University of Minnesota.

between the parenchymal cells, anastomose at the periphery of the lobule, and enter the bile ducts. The parenchyma is best defined as a continuous mass of cells perforated by a network of tunnels in which the sinusoids run, and the cells are called parenchymal cells. The walls of the sinusoids are lined by large cells with bulging nuclei, which are named Kupffer cells, and are functional phagocytes. The parenchymal cells, or hepatocytes, are the principal functioning cells of the liver.

General Functions of the Liver

Several general functions of liver are: to furnish the metabolic fuel for a variety of tissues in fasting, that is, glucose, amino acids, and ketone bodies; to remove from the blood metabolic waste products generated by other tissues for conversion to other metabolic fuels or for conversion to compounds for excretion via the urine or feces; to synthesize compounds used as structural components for other tissues; and to detoxify biologically active compounds such as drugs, hormones, and poisons.

Metabolism

Carbohydrate Metabolism

A major function of liver is blood glucose homeostasis. The liver converts glucose in the blood to glycogen after a meal and converts glycogen to glucose in fasting, and also converts amino acids and lactate to glucose in fasting.

The liver parenchymal cells, in contrast to adipose and muscle cells, do not have a permeability barrier to glucose; thus insulin has no effect on uptake. Insulin, however, does stimulate the metabolism of glucose by liver, by stimulating glucokinase activity. Liver hexokinase and glucose 6-phosphatase also catalyze the formation of glucose 6-phosphate from glucose. Glucose 6-phosphatase is an enzyme that has two different catalytic functions. Its hydrolytic reaction is as follows:

$$\text{Glucose 6-phosphate} + H_2O \longrightarrow \text{glucose} + P_i$$

The enzyme also catalyzes the following:

$$\text{Glucose} + \text{pyrophosphate } (PP_i) \longrightarrow \text{glucose 6-phosphate} + \text{Pi}$$

The major fate of glucose 6-phosphate in liver is conversion to glycogen. Small amounts of glucose 6-phosphate enter the glycolytic pathway and the hexose monophosphate shunt pathway for the pur-

poses of fatty acid synthesis. The control step for glycogen synthesis in liver is glycogen synthase, which is activated by a dephosphorylation by a phosphatase.

The breakdown of glycogen to glucose 1-phosphate by phosphorylase in fasting is under the control of glucagon and epinephrine. The effects of epinephrine on liver glycogen breakdown are mediated by α-adrenergic receptors and involve a *cAMP-independent* mechanism, *and* also by β-adrenergic receptors involving the production of cAMP. The stimulation of α-adrenergic receptors is postulated to mobilize intracellular Ca^{2+}, which in turn stimulates phosphorylase kinase. The effect of glucagon on liver glycogen breakdown is mediated by β-adrenergic receptors and involves cAMP-dependent mechanisms and also guanyl nucleotides.

The amount of glycogen stored in liver is not sufficient to maintain blood glucose levels in fasting, especially for periods of time exceeding about 5 h after a meal. The major source of blood glucose in fasting is gluconeogenesis in liver, which, like glycogen breakdown, is stimulated by the hormones glucagon and epinephrine. The process of gluconeogenesis is stimulated by the glucocorticoids in prolonged fasting and is inhibited by insulin after a meal. The mechanism of the action of glucagon in gluconeogenesis has been investigated more thoroughly than the rest of the above listed hormones. In general, the action of glucagon is to stimulate or to inhibit key control steps in gluconeogenesis or glycolysis, respectively, by a variety of mechanisms. The reactions catalyzed by fructose 1,6-bisphosphatase and by phosphofructokinase are irreversible steps in gluconeogenesis and glycolysis, respectively. The activation of fructose 1,6-bisphosphatase and the inhibition of phosphofructokinase by glucagon is mediated via cAMP *and* by the presence of free fatty acids liberated from adipose tissue. The action of cAMP on phosphofructokinase may be to modify the catalytic properties of the enzyme directly *or* cAMP may act via a phosphorylation of the enzyme by a cAMP-dependent protein kinase. Glucagon stimulates the inactivation of the glycolytic enzyme pyruvate kinase by a cAMP-dependent phosphorylation. The reaction catalyzed by pyruvate carboxylase is an irreversible step in gluconeogenesis, and acetyl CoA is an allosteric activator of pyruvate carboxylase. The stimulation of the release of free fatty acids from adipose tissue by glucagon has been described, and their oxidation in liver results in an increase in the acetyl CoA levels in the liver mitochondria, which in turn activates the pyruvate carboxylase. The reaction catalyzed by phospho*enol*pyruvate carboxykinase is markedly increased in fasting by increases in the amount of enzyme protein. The stimulatory action of glucagon on the catabolism of branched-chain amino acids in skeletal and cardiac muscle has been described (page 1012). One

consequence of this action in muscle is to increase the production and release of alanine and glutamic acid to the blood for transport to the liver and conversion to glucose. In summary, glucagon not only acts in a variety of ways on key control steps to inhibit glycolysis and to stimulate gluconeogenesis, but it is responsible for the increased production of substrates to be converted to glucose in the liver.

The mechanisms of the stimulation of gluconeogenesis by epinephrine are less well understood, and seem not to involve cAMP directly, but rather epinephrine is thought to act via Ca^{2+}. The gluconeogenic effects of epinephrine could be partially explained by its lipolytic effect on adipose tissue and its stimulatory effects on branched-chain amino acid catabolism in muscle. That is, the epinephrine-induced release of fatty acids from adipose tissue results in an increased fatty acid oxidation and acetyl CoA production in liver mitochondria, with an activation of pyruvate carboxylase by the acetyl CoA. The epinephrine-induced stimulation of branched-chain amino acid catabolism in heart and skeletal muscle results in an increase in the transport of the amino acids alanine and glutamate to the liver for conversion to glucose. The stimulation of gluconeogenesis by the glucocorticoids is thought to be by the stimulation of the breakdown of tissue proteins which are transported to the liver for conversion to glucose. Three mechanisms have been proposed for the inhibitory action of insulin on gluconeogenesis: an inhibition of the hepatic uptake of compounds that the liver converts to glucose; the stimulation of the incorporation of amino acids into tissue protein; and the antilipolytic effect of insulin on adipose tissue (page 774).

Approximately 15% of the glucose metabolized by the liver is by the hexose monophosphate shunt for the purpose of generating NADPH for fatty acid and cholesterol synthesis. The rate of metabolism of glucose by the hexose monophosphate shunt is increased in feeding and decreased in fasting. The activity of glucose 6-phosphate dehydrogenase increases about fivefold after feeding, and the increase in enzyme activity has been accounted for by increases in the amount of enzyme. In addition, a high activity form of glucose 6-phosphate dehydrogenase has been found in livers of fed rats.

The amount of glucose metabolized by glycolysis in liver is small when compared to other tissues such as muscle or brain. The main reason for glycolysis in liver is to furnish pyruvate for conversion to acetyl CoA to be used for fatty acid synthesis. The rate of glycolysis in liver is increased in feeding and decreased in fasting, like the hexose monophosphate shunt.

Fructose is also metabolized by the liver and two pathological conditions occur leading to fructosuria (Clin. Corr. 21.2).

CLIN. CORR. 21.2 ESSENTIAL FRUCTOSURIA

Hepatic fructokinase is deficient in patients with fructosuria, and these individuals lack the ability to phosphorylate fructose to fructose 1-phosphate. Patients with hereditary fructose intolerance lack the hepatic enzyme fructose 1-phosphate aldolase.

Lipid Metabolism

Liver from fed animals will synthesize fatty acids from glucose and also will incorporate fatty acids from plasma triglycerides, and the fatty acids are, in both cases, used for the synthesis of phospholipids. The phospholipids are transported as lipoproteins in the blood to various other tissues. Liver from fasted animals will utilize free fatty acids, released from adipose tissue, as its metabolic fuel. The activities of all three control enzymes in fatty acid synthesis (citrate cleavage enzyme, acetyl CoA carboxylase, and fatty acid synthetase) are higher in livers from fed than from fasted rats. Acetyl CoA carboxylase is activated (page 446) by the efflux of citrate from the mitochondria in feeding, and by a dephosphorylation mediated in some way by insulin. The inactivation of acetyl CoA carboxylase in fasting is mediated by glucagon and by a phosphorylation reaction. The increase in fatty acid synthetase in feeding is accounted for by an increase in enzyme protein. The carnitine transport system for entry of fatty acids into the mitochondria (page 468) is the control step that results in fatty acid oxidation by mitochondria in fasting or incorporation of fatty acids into phospholipids in feeding.

Protein Metabolism

Rapid protein synthesis and breakdown occurs in liver. The only serum proteins not synthesized in liver are the γ-globulins. Plasma proteins are constantly removed from circulation by liver and hydrolyzed to amino acids for utilization by extrahepatic tissues. The proteins enter the cells by a process of pinocytosis and are degraded to amino acids by intracellular cathepsins and other proteolytic enzymes in the lysosomes. The half-life of both serum and liver protein is about 10 days in contrast to the half-life of muscle protein which is much slower, that is, about 180 days.

The Metabolic Fuel of Liver

The energy demands of the liver are large and are supplied by lactate after a meal and fatty acids released from adipose tissue in fasting. Liver does have lipoprotein lipase activity, and therefore plasma triglycerides are a possible source of fatty acids for the organ in fasting.

Biotransformation Reactions

A variety of biotransformation reactions occur solely or mainly in liver which are essential for the maintenance of the entire animal. A number of these transformations have been described in detail and will only be listed here: synthesis of bile acids from cholesterol (page

514); reduction of adrenal steroids and conjugation with glucuronic acid or sulfuric acid (page 729); conjugation of androgens and estrogens with glucuronic acid or sulfuric acid (page 742); formation of bile pigments (page 1164); storage of iron (page 1069); synthesis of purine and pyrimidine bases (page 637); formation of uric acid (page 649); and synthesis of urea (page 551). The pathway for creatine synthesis is shown in Figure 21.14.

The liver detoxifies a variety of drugs and poisons by a number of reactions: oxidation, reduction, hydrolysis, conjugation, and methylation.

Oxidative Reactions

Enzymes catalyzing oxidation reactions are present in the microsomal fractions of liver. The enzymes require NADPH and molecular oxygen and catalyze a variety of oxidative reactions commonly called "hydroxylation" reactions. Included in these reactions are aromatic and aliphatic hydroxylation, *N*-, *O*-, and *S*-dealkylations, sulfoxidation, *N*-oxidation, and epoxidation. The microsomal

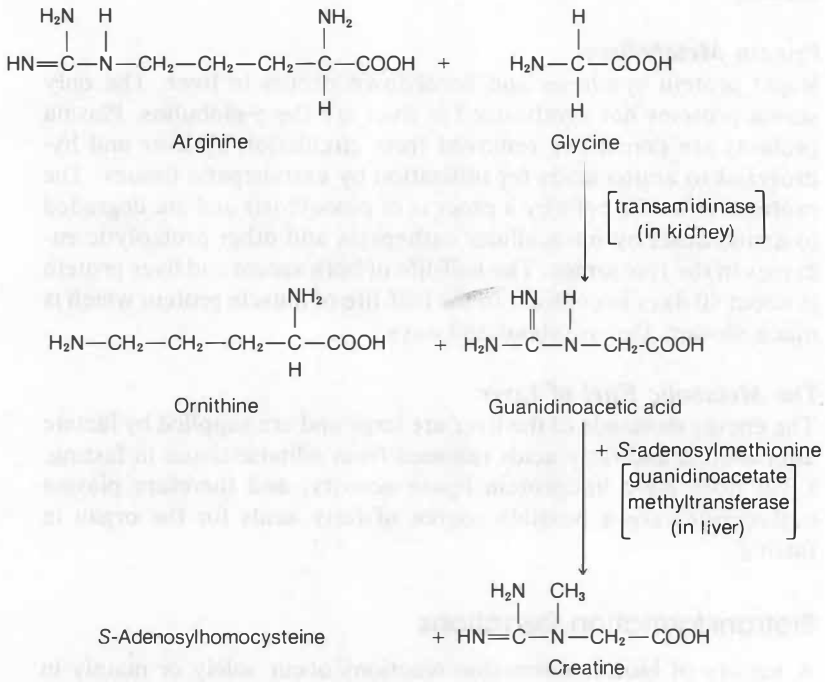
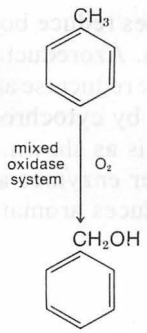


Figure 21.14
Biosynthesis of creatine.

oxidase systems, commonly known as mixed function oxidases, are comprised of three components: a hemoprotein (called cytochrome P_{450}), a flavoprotein reductase and a lipid factor. A list of some compounds oxidized by this enzyme system is shown in Table 21.4. The general reaction catalyzed by mixed function oxidases is as follows:

$$RH + O_2 + NADPH \longrightarrow ROH + H_2O + NADP^+$$

A specific reaction is the oxidation of toluene. The administration of a large number of compounds induce liver mixed function oxidase activity. The compounds noted for their induction of this enzyme activity are the barbiturates and polycyclic hydrocarbons such as benzo(*a*)pyrene.



The oxidation of alcohols to aldehydes or ketones and the oxidation of aldehydes to carboxylic acids are catalyzed by two groups of enzymes in liver that do not involve the cytochrome P_{450} system: (1) The pyridine nucleotide-linked oxidoreductases utilize NAD^+ and are located in both the cytosol and mitochondria; (2) aldehyde oxidase, a metalloflavoprotein. The best known example of NAD^+ linked oxidoreductases is liver alcohol dehydrogenase, which catalyzes the following:

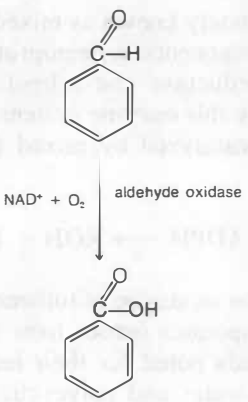
$$CH_3CH_2OH + NAD^+ \longrightarrow CH_3\overset{\overset{O}{\|}}{C}-H + NADH$$

Ethanol — Acetaldehyde

Liver aldehyde dehydrogenase catalyzes the oxidation of a number of aldehydes, that is, formaldehyde and acetaldehyde, to their alcohols. Aldehyde oxidase is a metalloflavoprotein containing iron, molybdenum, and FAD, and catalyzes the oxidation of benzaldehyde to benzoic acid.

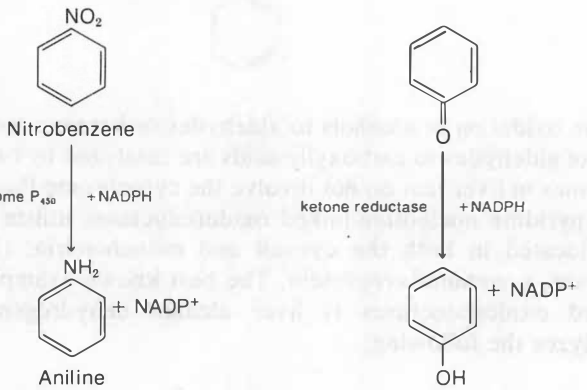
Table 21.4 Compounds That Are Oxidized by Mixed Function Oxidase in Liver

3-Methyl-4-aminoazobenzene
Aminopyrine
Biphenyl
Aniline
Dichloromethane
Benzo(*a*)pyrene
7-Ethoxycoumarin
N-Methyl-*p*-chloroaniline
Ethylmorphine



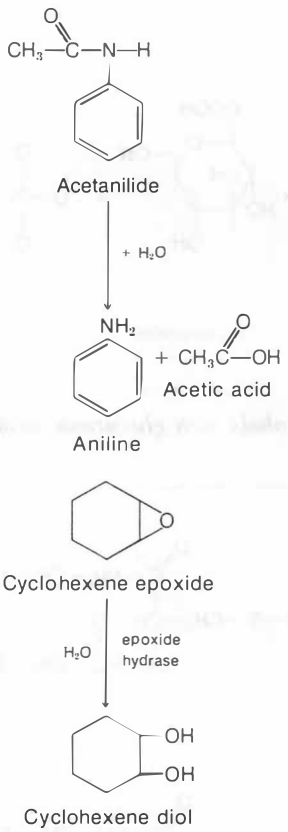
Reductive Reactions

Hepatic microsomal enzymes reduce both azo and nitro compounds by the addition of hydrogen. Azoreductase activity is attributable to both NADPH–cytochrome c reductase and cytochrome P_{450}, whereas nitroreductase is catalyzed by cytochrome P_{450} only. The reduction of nitrobenzyme to aniline is as shown. Aldehydes and ketones are reduced to alcohols by liver enzymes and the coenzyme NADPH. Liver ketone reductase reduces aromatic ketones to their alcohols.



Hydrolytic Reactions

Esterases are found in a variety of tissues, for example, plasma, brain, intestinal mucosa, erythrocyte, and muscle. An esterase is found in human liver that catalyzes the hydrolysis of acetanilide, procaine, xylocaine, and simple aliphatic esters. Epoxide hydrase catalyzes the conversion of epoxides to diols.



Conjugation Reactions

A large number of compounds are detoxified by conjugation reactions in the liver. Four types of compounds are conjugated with *glucuronic acid:* hydroxyl (both phenolic and alcoholic), carboxyl, sulfhydryl, and amino. The enzyme catalyzing the formation of glucuronides, called UDP-glucuronyltransferase, is found in the endoplasmic reticulum and a reaction it catalyzes is shown in Figure 21.15. Glucuronide formation is a major pathway of drug metabolism. Compounds bearing alcoholic or phenolic hydroxyl groups are conjugated with *sulfate,* catalyzed by the enzyme sulfotransferase, found in the cytosol of the liver cell. The sulfate group is transferred from 3′-phosphoadenosine 5′-phosphosulfate (PAPS) to yield the sulfate half ester as follows:

$$ROH + PAPS \longrightarrow ROSO_3H + PAP$$

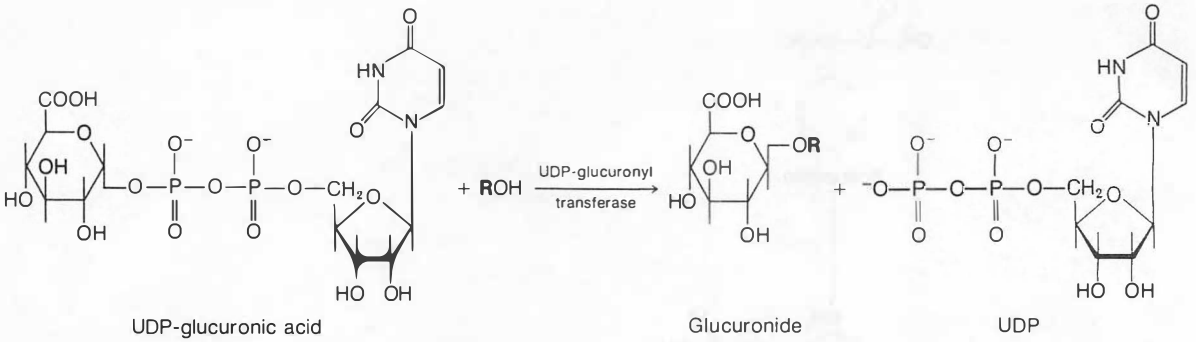


Figure 21.15
The conjugation of alcohols with glucuronic acid in liver.

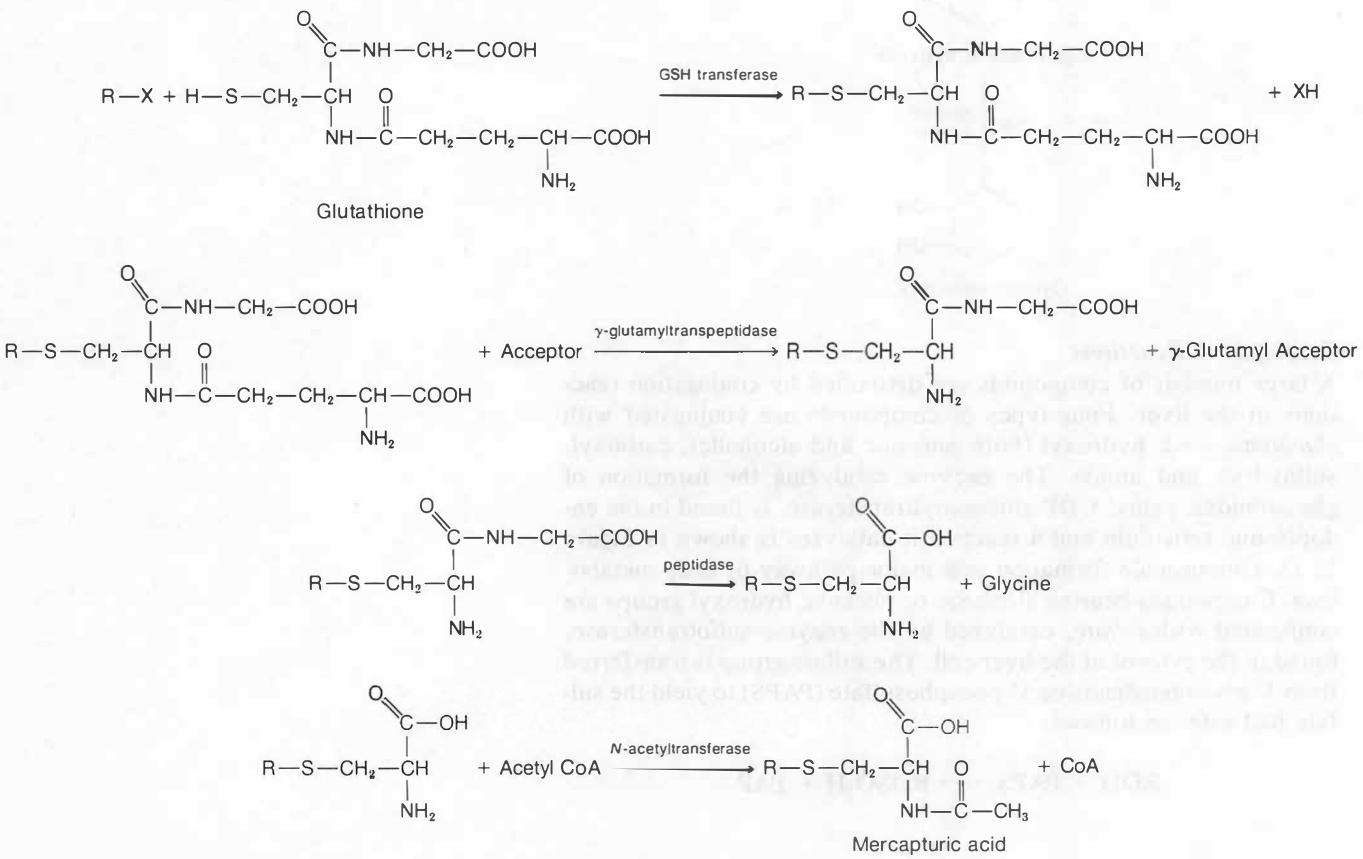
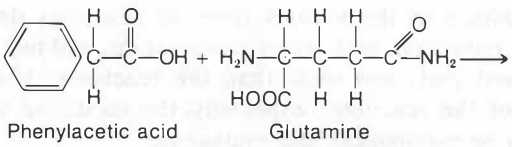
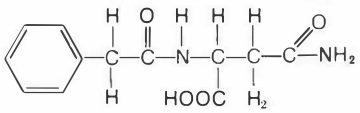


Figure 21.16
The pathway of the formation of mercapturic acids in liver.

Glutathione conjugates are thioethers formed by the combination of alkyl and aryl halides, epoxides, and alkenes with glutathione. The enzymes that catalyze these reactions are glutathione transferases, found in the cytosol. Some of glutathione conjugates are converted to mercapturic acids, catalyzed by *S*- and *N*-acetyltransferases. The formation of mercapturic acids is shown in Figure 21.16. *Glutamine* is conjugated with phenylacetic acid, and a number of drugs. The conjugation of phenylacetic acid with glutamine is as follows:

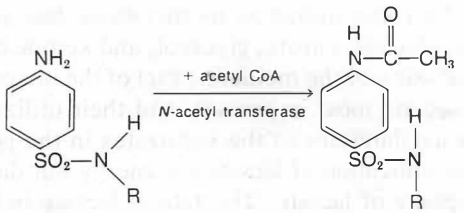


Phenylacetic acid Glutamine

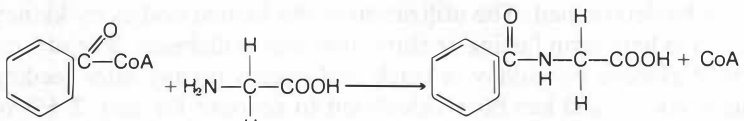


Phenacetylglutamine

Sulfanilamide is detoxified by forming the conjugate with *acetic acid.*



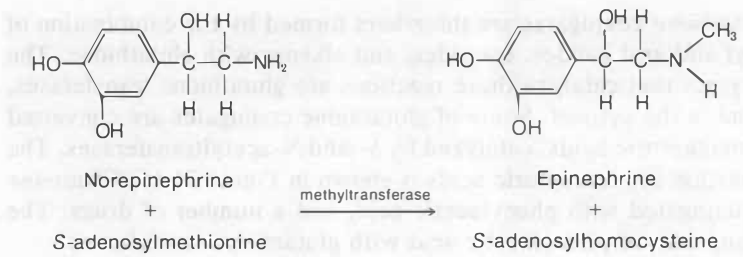
Benzoic acid is conjugated with *glycine* to form hippuric acid.



benzoyl CoA Glycine Hippuric acid

Methylation Reactions

A variety of methyltransferases that utilize *S*-adenosylmethionine as the methyl donor are found in liver. For example, catechol-*O*-methyltransferase catalyzes the *O*-methylation of norepinephrine, dopamine, epinephrine, dopa, and a number of drugs. See also the *N*-methylation of norepinephrine to epinephrine.



The products of the various types of reactions described (i.e., oxidation, reduction, hydrolysis, conjugation, and methylation) are, for the most part, less toxic than the reactants. However, some products of the reactions, especially the oxidative reactions, are thought to be carcinogens and mutagens.

21.5 KIDNEY

Metabolic Fuel of Kidney

Mammalian kidney can utilize as its metabolic fuel palmitate, lactate, glutamine, glucose, citrate, glycerol, and ketone bodies. Palmitate furnishes 60–80% of the metabolic fuel of the intact kidney, with lactate as the second most important, and their utilization depends on the relative availabilities of the substrates in the plasma. Palmitate inhibits the utilization of lactate for energy but does not inhibit the kidney's uptake of lactate. The fate of lactate in kidney in the presence of large amounts of palmitate is conversion to glucose by gluconeogenesis. The relative contributions of citrate, glutamine, and glycerol to the energy requirements of the intact kidney have as yet to be determined. The utilization of the ketone bodies by kidney occurs in long-term fasting or starvation and in diabetes. The utilization of glucose by kidney is small and occurs mainly after feeding and in anoxia and has been calculated to account for only 2–6% of the oxygen consumption of the intact kidney. The major source of energy for the production of ATP by kidney is aerobic oxidation of substrates, with glycolysis producing only 4% of the ATP. The major utilization of ATP in kidney is for the reabsorption of NaCl by the kidney tubules.

Kidney cortex and kidney medulla differ greatly in their metabolism. Kidney cortex has a high rate of oxygen consumption, large amounts of citric acid cycle and respiratory chain enzymes, and a respiratory quotient of ~0.75 indicating that fatty acids are the

principal metabolic fuel. The utilization of palmitate, lactate, ketone bodies, and glucose by oxidative metabolism in the intact kidney for energy occurs mainly in the kidney cortex. The oxidation of fatty acids by the cortex is accompanied by ketone body formation, as it is in liver. The medulla of the kidney has, on the other hand, high concentrations of the enzymes of glycolysis, low levels of the citric acid cycle and respiratory chain enzymes, low rates of oxygen consumption, and low energy requirements, which arise primarily from glycolysis. The source of the glucose for glycolysis by kidney medulla is plasma glucose after a meal and, perhaps, glucose synthesized in the cortex in fasting. The high respiratory metabolism of the cortex and the low respiratory metabolism of the medulla is correlated with a Po_2 of 80 mm, and 5–10 mm, in the cortex and medulla, respectively.

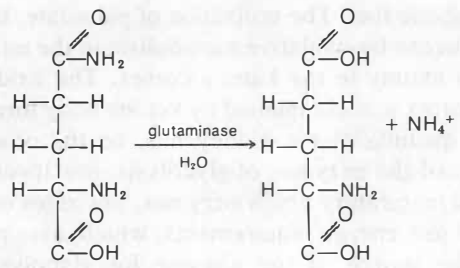
Gluconeogenesis in Kidney Cortex

The gluconeogenic enzymes are present in kidney cortex and kidney cortex can synthesize glucose more rapidly per unit weight than can liver, but the kidney does not contribute significant amounts of glucose to the bloodstream. The suggestion has been made that the glucose synthesized in the cortex is used by the medulla as its metabolic fuel. The rate of gluconeogenesis in kidney cortex, like liver, is higher in fasting than in the fed state, and the hormones epinephrine and glucagon are implicated in the activation of the process via cAMP. The fatty acids and ketone bodies supply the energy for kidney cortex gluconeogenesis in fasting.

The rates of gluconeogenesis in kidney cortex are also stimulated in acidosis, but the physiological significance of this response to acidosis is not clear. The gluconeogenic enzyme phospho*enol*pyruvate carboxykinase is a major factor in the regulation of kidney cortex gluconeogenesis and the induction of its activites in acidosis is accompanied by increases in the amounts of enzyme present and by increases in the amount of messenger RNA for the enzyme. The nature of the signal acting on the renal cell to produce induction of the mRNA for PEP carboxykinase after acidosis is not known.

Ammonia Production in Kidney

Increased production and excretion of ammonium ion is a major component of the kidney's homeostatic response to onset of metabolic acidosis. Renal ammoniagenesis is initiated primarily by a phosphate-dependent glutaminase, which is localized within the inner membrane or matrix compartment of the mitochondria.



Glutamine — Glutamic acid

There is a specific carrier system for transport of glutamine across the inner mitochondrial membrane. Several factors have been proposed for the stimulation of renal ammonia formation in acidosis, including an increase in the activity or amount of the major ammonia-producing enzyme, glutaminase, an increased utilization of the end product inhibitor of this enzyme, glutamate, and an increase in the rate of glutamine transport into the mitochondria.

Kidney Tubular Transport Mechanisms

Kidney is one of the richest sources of Na^+, K^+-stimulated ATPase, which is present in very high concentrations in the thick ascending loop of Henle and the proximal tubules. The reaction catalyzed by this enzyme is as follows:

$$\underset{\text{(inside)}}{3Na^+} + \underset{\text{(outside)}}{2K^+} + ATP + H_2O \longrightarrow \underset{\text{(outside)}}{3Na^+} + \underset{\text{(inside)}}{2K^+} + P_i + ADP + H^+$$

The function of this enzyme is the active transport of K^+ and Na^+ into and outside of the cell, respectively. An Na^+, K^+-ATPase has been isolated in a pure form from renal medulla and when constituted into phospholipid vesicles will exhibit active K^+ transport coupled to active Na^+ countertransport in a 2 : 3 ratio. The primary overall ion movement occurring in the ascending loop of Henle is, however, reabsorption of NaCl from the lumen of this tubular structure into peritubular space. An ATP-dependent (active transport) of NaCl has been identified in the ascending loop of Henle. Any modification of the function of Na^+, K^+-ATPase to produce net NaCl reabsorption (in addition to its function as a K^+-Na^+ exchange) in the kidney is thought to be due to the presence of kidney factors (perhaps proteins) that are absent in nervous tissue and red blood cells.

The transport of both D-glucose and amino acids (in the proximal

region of the nephron) from the glomerular filtrate into the blood is energy-dependent, saturable, and is stimulated by Na^+ (or cotransported with Na^+). Sugars and amino acids have been suggested to compete for the Na^+ for cotransport. The coupling of sugar and amino acid transport to the reabsorption of Na^+ may well be responsible for 10–15% of the volume uptake in the proximal tubule.

21.6 NERVOUS TISSUE

Structure

The major feature of the nerve cell, or neuron, is the presence of greatly elongated cytoplasmic processes, which in certain instances extend almost the length of the body. By means of these processes the neuron contacts other neurons, receptors, glands, or muscles and thus transmits or receives stimuli from them. There are about 10 billion neurons in the nervous system and most of these are located in the brain. All neurons have a cell body (a swollen portion of the neuron), which contains abundant cytoplasm and a large centrally located nucleus. The long extension of the cytoplasm that conducts the impulse away from the cell body is called the axon. The short aborized extensions of the cytoplasm that receive impulses from other neurons are called dendrites or dendrons. A motor nerve cell and investing membranes is depicted in Figure 21.17. The neurons are protected by various types of tubular structures (e.g., myelin sheath) that are not considered part of the neurons (see Clin. Corr. 21.3). A schematic diagram of an axon surrounded by myelin sheath is shown in Figure 21.18. The composition of the cytoplasm in the cell body and the extensions of the cytoplasm are not identical. The cytoplasm of the cell body contains Golgi apparatus, mitochondria, pigment granules (melanin and lipofusin), droplets of lipid and glycogen, neurofibrils, neurotubles, neurofilaments, microfilaments, and rough endoplasmic reticulum referred to as Nissl bodies. The cytoplasm of the axon contains neurofibrils and mitochondria, but no rough endoplasmic reticulum, pigment, or lipid droplets. The absence of rough endoplasmic reticulum in the terminal regions of the dendrites and in the axons means that proteins are synthesized in the cell body and are transported down the interior of the axons and dendrites. Gray matter is the core of the nervous system containing cell bodies of the neurons, their dendrites, and the proximal portions of the axons. The zone mainly devoid of cell bodies and heavily invested with myelin is white matter.

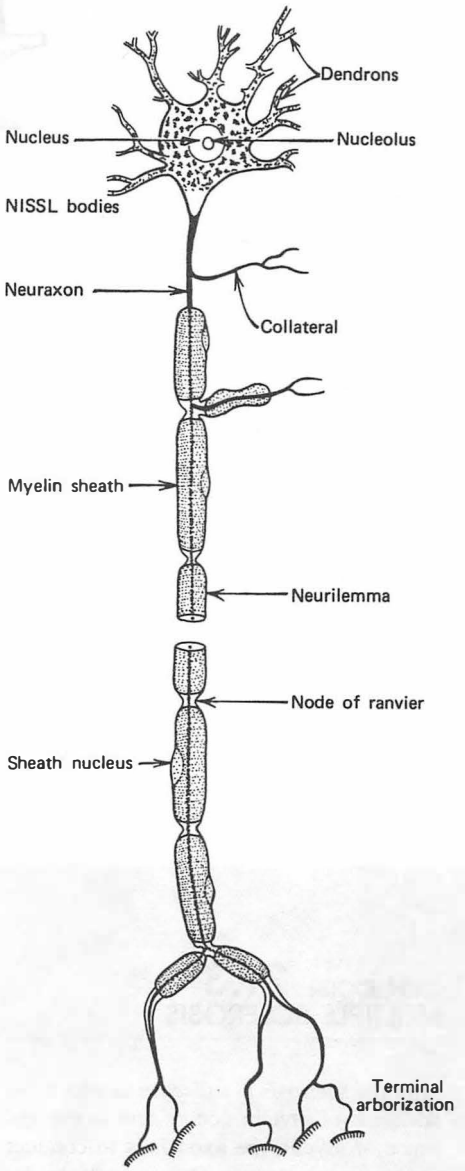


Figure 21.17
A motor nerve cell and investing membranes.

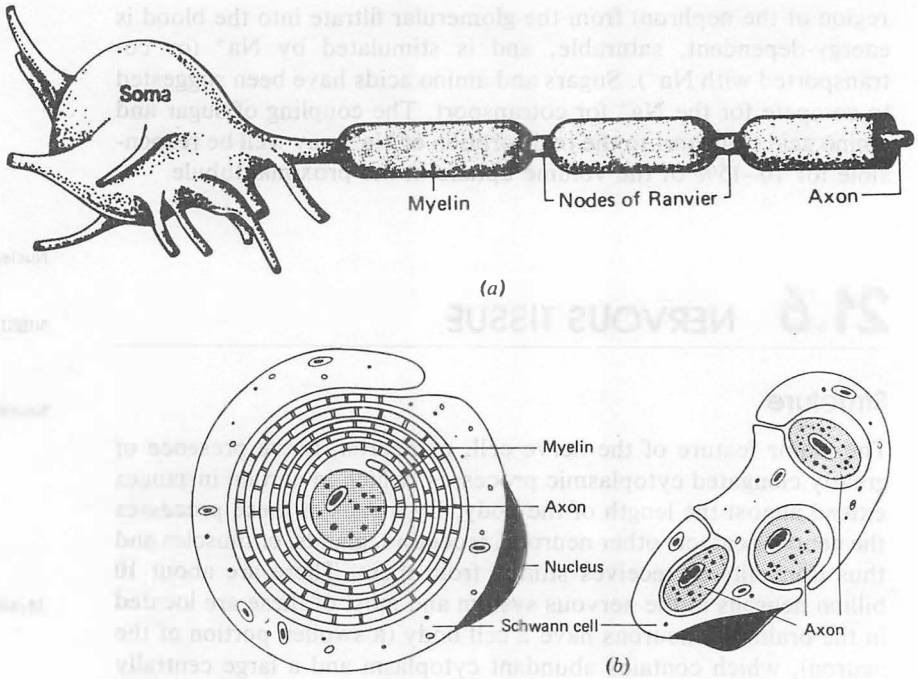


Figure 21.18
A schematic diagram of axon surrounded by myelin sheath and the unmyelinated node of Ranvier.
The drawing in the lower panel (b) is a cross section of a nerve, showing the axon surrounded by the myelin sheath composed of layers of Schwann cell plasma membrane.
Reprinted with permission from R. F. Schmidt, ed., *Fundamentals of Neurophysiology*, Springer-Verlag, New York, 1975, p. 8.

CLIN. CORR. 21.3
MULTIPLE SCLEROSIS

Multiple sclerosis is a disease in which destruction of myelin occurs and in the absence of myelin the axon fails to conduct the nervous impulses. The cause of the destruction of myelin is uncertain, but a genetic predisposition for the disease occurs.

Nervous Tissue Lipids

A major characteristic of nervous tissue is its high content of lipid and their unique structures. The lipid content of myelin, white matter, and gray matter is 80, 60, and 40% dry weight, respectively. Cholesterol is the second most abundant substance in brain, and about 25% of the body cholesterol is present in nervous tissue. Cholesterol is synthesized in the brain and has an extremely low rate of breakdown or turnover number. Water is the most abundant substance in brain, and the water content, expressed as percent of fresh weight of myelin, white matter, and gray matter are 40, 70, and 80% of fresh weight, respectively.

Most of the fatty acids in the brain lipids are synthesized in situ,

with only small amounts incorporated from the diet. The fatty acid composition of brain lipids is relatively constant, but variations have been reported with alterations in the lipid composition of the diet. Many of the fatty acids in nervous tissue are unsaturated and have long carbon chains, as high as 24 carbons. The palmitate, synthesized in the cytoplasm, is elongated by a mitochondrial and microsomal system, utilizing acetyl and malonyl CoA, respectively. Two enzymatic systems have been identified that desaturate the carbon chains between carbons 6-7 and 9-10, respectively.

The lipids found in nervous tissue and the locations where they are found in relatively high concentrations are shown in Table 21.5. All of the lipids listed in this table are found in varying amounts in white and gray matter and in myelin. The pathways for the synthesis and breakdown of the lipids found in nervous and other mammalian tissues are described in Chapter 10. The source of choline and serine for brain phosphatidylcholine and phosphatidylserine synthesis, respectively, is exogenous (i.e., abstracted from the blood). The source of ethanolamine and inositol for the synthesis of phosphatidylethanolamine and phosphatidylinositols, respectively, are both endogenous and exogenous. No metabolic role has been assigned to any of the lipids found in nervous tissue other than that of cellular membrane functions.

Nervous Tissue Proteins

Attempts to relate the structure of major proteins found in nervous tissue to the function of nervous tissue have not as yet been success-

Table 21.5 The Cerebral Lipids and the Sites in Which They Are Found in Relatively High Concentrations

Cholesterol	Approximately equal in myelin, white matter and gray matter
Phosphatidylcholine Phosphatidylethanolamine Phosphatidylserine	Gray matter
Phosphatidylinositols	Myelin (peripheral nerve)
Spingomyelin	Approximately equal in myelin, white matter, and gray matter
Cerebrosides	Myelin and white matter
Sulfatides	Myelin and white matter
Gangliosides	Gray matter
Plasmalogens	Myelin and white matter
Polyglycerophosphatides	Gray matter

ful. Three proteins have received some attention. A highly acidic protein termed S-100 is found in large amounts in the glial cells of the brain and also in small amounts in neurons. Glutamic and aspartic acid comprise 30% of the amino acid residues. The protein is comprised of three-nonidentical subunits, and it has a high affinity for Ca^{2+}. The molecule undergoes a conformational change in the presence of Ca^{2+}. Another protein, given the designation of 14-3-2 from its sequence of elution in chromatographic steps is found mainly in neurons, and is also a highly acidic protein. The 14-3-2 protein has now been identified as an isozyme of enolase, the enzyme that catalyzes the reversible conversion of 2-phosphoglycerate to phospho*enol*pyruvate. Both the S-100 and the 14-3-2 protein have been implicated by some investigators with the memory or learning ability of the brain. Calmodulin, necessary for the Ca^{2+} activation of the enzyme cyclic nucleotide phosphodiesterase, is found in brain, and also is thought to be involved in the Ca^{2+}-dependent release of acetylcholine and norepinephrine from their vesicular stores.

$$\text{Norepinephrine (in vesicles in cytoplasm of axon)} \xrightarrow[\text{calmodulin}]{Ca^{2+}} \text{norepinephrine (discharged from axon)}$$

$$\text{cAMP} \xrightarrow[\text{calmodulin phosphodiesterase}]{Ca^{2+}} \text{5'-AMP}$$

The major proteins found in myelin are a *basic protein* and a *proteolipid*. The basic protein, mol wt 18,000, constitutes 30% of the total protein in myelin, and contains a large number of arginine residues. Proteolipid is a combination of a protein, composed of a large number of hydrophobic amino acids, with a lipid. Proteolipid and the basic protein are located on the outer and inner surfaces of the myelin membrane, respectively, and abut each other within the lipid bilayer. A schematic representation of the arrangement of myelin basic protein and proteolipid in the myelin membrane is shown in Figure 21.19.

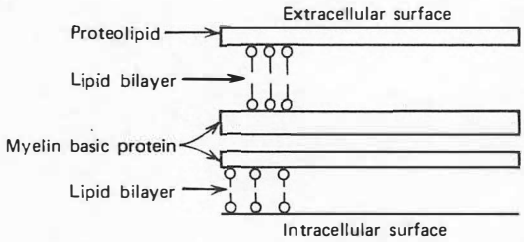


Figure 21.19
Schematic representation of the arrangement of myelin basic protein and proteolipid in the myelin membrane.

Neurotubular, Neurofilament, and Microfilament Protein

The slender tubules transversing the cytoplasm from one dendrite to another or to an axon have a role in axonal flow or secretion. Microtubules are also probably involved in cell differentiation, intracellular transmission of signals from cell surface receptors and the transport of effectors of enzymes. Microtubules are involved in the regulation of cholesterol synthesis in glial cells. Tubulin is the major protein found in microtubules and comprises about 14% of the total protein in mammalian brain. Tubulin consists of two similar but not

identical subunits, mol wt 55,000–60,000. Purified tubulin in the presence of Mg^{2+} and GTP will reconstitute or reassemble into microtubules. A diagrammatic view of the arrangement of the subunits of tubulin in the microtubules is shown in Figure 21.20. The most prominent protein in the neurofilaments is a component chemically distinct from tubulin. The neurofibrils are bundles of neurofilaments, and the protein of the microfilaments is actin.

The brain proteins have a rapid turnover rate relative to other body proteins. A mean half-life of brain proteins has been calculated to be about 85 h, or about 20 times faster than total body protein. Liver or serum proteins turnover at one-third the rate of brain protein. Data on the constant amounts of protein in brain, the high proteolytic activity of nervous tissue, the rapid incorporation of glucose into brain proteins, and the small arterial–venous differences of nitrogenous compounds for brain have been interpreted to indicate a retention and reutilization of NH_4^+ by nervous tissue, which is unique from any other mammalian tissue. The urea cycle enzymes are present in nervous tissue for, as yet, some unexplained reason. A retention and reutilization of the breakdown products of brain phospholipids is also known to occur. The retention and reutilization of brain protein and lipid seems logical in view of the relative impermeability of the "blood–brain barrier" and the resistance of the brain to deterioration during long-term starvation. The brain is the only tissue in which the breakdown products of protein and phospholipid are extensively reutilized. The other individual tissues in the body release their cellular breakdown products to the circulatory system for processing by the liver.

The Action Potential

Na^+ and K^+ are involved in the excitation and transmission of nervous impulses. The potential difference between the interior and exterior of a resting nerve cell is about 60 mV with the inside negative and the outside positive. The potential difference is mainly the result of the large amounts of K^+ on the inside of the cell relative to the outside of the cell. The sodium concentration outside the cell is larger than it is inside the cell, but this difference does not contribute greatly to the resting potential. The K^+ and Na^+ gradients are maintained by the Na^+, K^+-ATPase, whose action is to pump Na^+ out and K^+ into the cell. If a nerve cell is stimulated, the membrane becomes depolarized, and the potential difference between the inside and outside of the cell changes, and this change is called the action potential. The first event that produces the action potential is the increase of the permeability of the membrane to Na^+. A small amount of Na^+ enters the cell and the inside of the cell becomes positive with re-

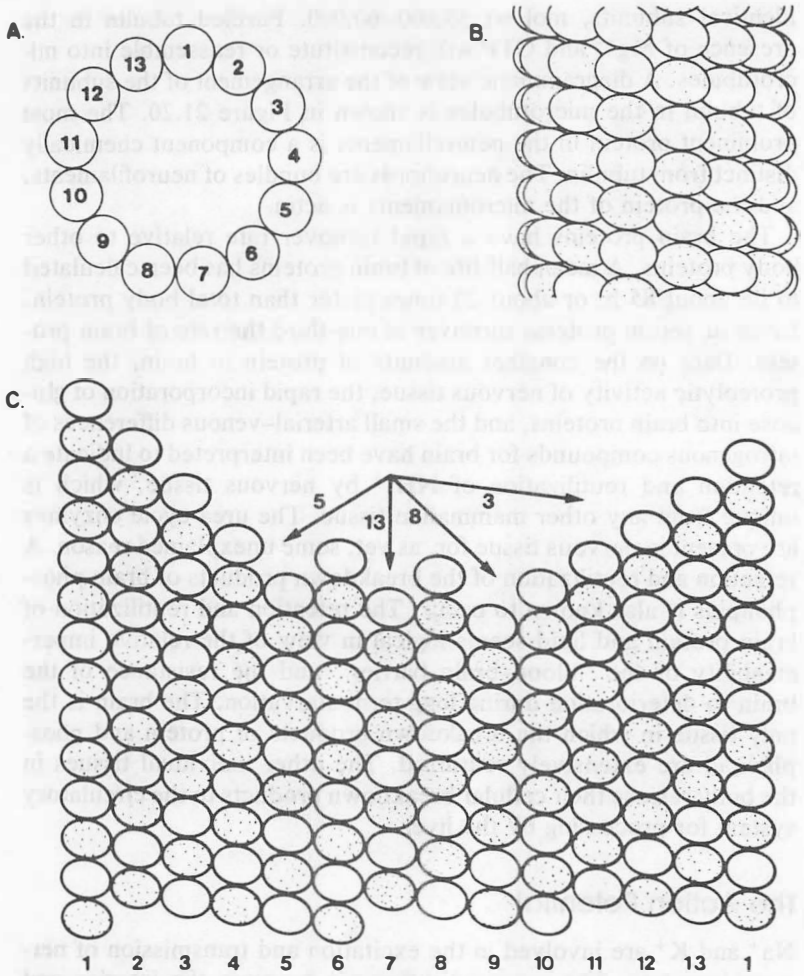


Figure 21.20
The arrangement of subunits of tubulin in microtubules.
(a) Diagramatic view of a cytoplasmic microtubule cross section showing arrangement of 13 protofilaments. (b) Diagrammatic view of cytoplasmic microtubules surface lattice showing the probable arrangement of the heterodimers. Stippled and clear ellipsoids represent α and β subunits. Monomers alternate along the three-start helix and within a protofilament. (c) Representation of the helix net as seen in an unfolded microtubule. Numbers at the bottom of the sheet correspond to the numbered protofilaments in (a). Each ellipsoid represents a monomer, stippled and clear ellipsoids representing the α and β subunits, respectively. Subunits alternate along the three-start helix, and the α–β dimer axis is parallel to the

spect to the outside; that is, the membrane potential difference is about +35 mV. The potential difference is restored to the resting potential (~60 mV) by an increase in the permeability of the membrane to K^+, which leaks out because of the high intracellular K^+ concentration, and by a decrease in the rate of Na^+ entering the cell. These events are depicted in Figure 21.21. The potential difference is thus restored to its resting level, but the gradients between the intracellular and extracellular K^+ and Na^+ concentrations are not those found in the resting nerve. The ionic gradients are restored to the resting level by the pumping of Na^+ out and K^+ into the cell. This occurs with an expenditure of energy by the Na^+,K^+-ATPase. Mammalian brain, like heart and skeletal muscle, has a reservoir of high energy phosphate in the form of phosphocreatine with which the ATP supplies are sufficient to maintain the brain for 15–20 s.

Metabolism of Nervous Tissue

Carbohydrate

The brain of a well-nourished animal utilizes large amounts of glucose as its only metabolic fuel. Glucose is nearly completely oxidized to CO_2 and H_2O: actually 25% of the oxygen consumption of the adult body is due to glucose oxidation in brain. The oxygen consumption of brain in infants and children up to 4 years of age accounts for about 50% of the total body's oxygen consumption. The glycolytic rate functions at about 20% of its capacity but the citric acid cycle works at near the maximum potential rate in a well-oxygenated brain, and small amounts of lactate leave the brain. The greatest danger to the survival of brain is anoxia and the brain's protective mechanism against anoxia is the Pasteur effect (page 353). The rates of glycolysis and lactate formation have been found to increase five- and eightfold, respectively, after only 1 min of anoxia. The increase of the glycolytic flux in anoxia is attributed to the activation of hexokinase and phosphofructokinase. The details of activation of these enzymes in anoxia are not well understood and may be associated with a transfer of the enzyme from the cytoplasm to the mitochondria. Phosphofructokinase is also activated by NH_4^+,

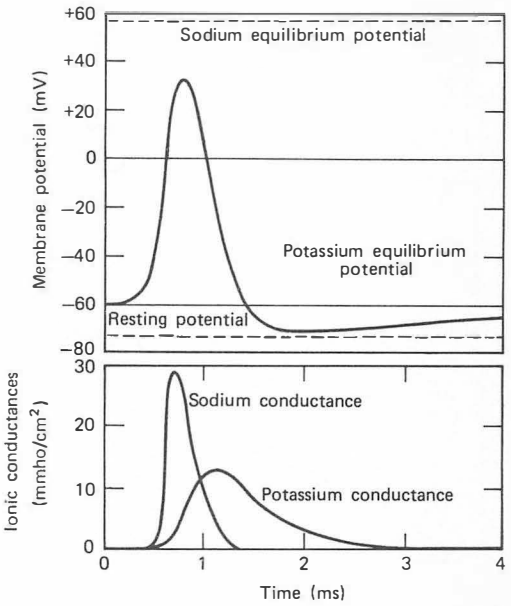


Figure 21.21
A schematic diagram of the potential changes in a neuron during stimulation.
Nerve impulse can be represented by changes in the voltage across the axon membrane (a) or by changes in the conductance of the membrane to sodium and potassium ions (b), both on a time scale of milliseconds. During the impulse the inside of the membrane becomes briefly positive with respect to the outside. After impulse has passed, the resting voltage is restored. From R. D. Keynes, Ion channels in the nerve cell membrane, *Sci. Amer.*, 240:126, 1979.

axis of the protofilament. The arrow designated 3 shows the angle of the three-start helix, arrow 13 the angle of the 13-start helix. α–β dimers can be viewed as arranged in a five-start right-handed helix (arrow 5) or an eight-start left-handed helix (arrow 8).
Reproduced with permission from J. A. Snyder and J. R. McIntosh, *Annu. Rev. Biochem.*, 45:706, 1976. Copyright 1976 by Annual Reviews Inc.

which is found to be increased in anoxia as compared to normal brain.

A deficiency of glucose in blood also endangers the survival of the brain. Endogenous glycogen can supply a small amount of glucose in the absence of blood glucose, but cerebral failure occurs quickly after sudden removal of glucose from the blood. Cerebral failure does not occur, however, with long-term hypoglycemia, such as in starvation, because ketone bodies replace glucose as the brain's metabolic fuel. The utilization of ketone bodies by human brain is thought to be due to their increased concentration in blood in starvation and not to an increase in the permeability of the blood–brain barrier or to an increase in the activities of the enzymes involved in ketone body utilization.

The control of the synthesis and breakdown of brain glycogen seems to be similar to that observed in other tissues. Both the D and I forms of glycogen synthase are found in brain along with the enzymes that catalyze their interconversions. Insulin is thought to have no effect on brain glucose or glycogen metabolism, probably because insulin does not cross the blood–brain barrier. About 3–5% of glucose metabolized in brain is by the hexose monophosphate shunt, to furnish NADPH for fatty acid and cholesterol synthesis.

Amino Acids

Human brain contains about eight times the amount of free amino acids found in plasma and the composition of the individual amino acids also varies greatly. The concentration of aspartic and glutamic acids are about 300-fold greater in brain than in plasma, and γ-aminobutyric acid, *N*-acetylaspartic acid, and glutamine are present in high amounts. The rates of uptake of amino acids from the blood by the brain are low relative to other tissues. Three different carrier systems for uptake into brain from blood for neutral, basic, and acidic amino acids have been found. The ability of the brain to retain and reutilize nitrogen for amino acid synthesis may well be one reason for the low uptake of blood amino acids into the brain.

Brain levels of NH_4^+ increase in anoxia or ischemia and the source of the NH_4^+ is believed to be the purine nucleotide cycle enzymes (page 1014). The immediate fate of the NH_4^+ is its use in the conversion of glutamic acid to glutamine. The utilization of the nitrogen from glutamate and glutamine for the synthesis of the nonessential amino acids from glucose in the brain seems likely. Some of the glutamate is decarboxylated to form the neurotransmitter γ-aminobutyric acid (GABA). GABA, unlike other neurotransmitters that are resequestered into vesicles after their release, is catabolized via a process called the GABA shunt. The GABA shunt is outlined in Figure 21.22.

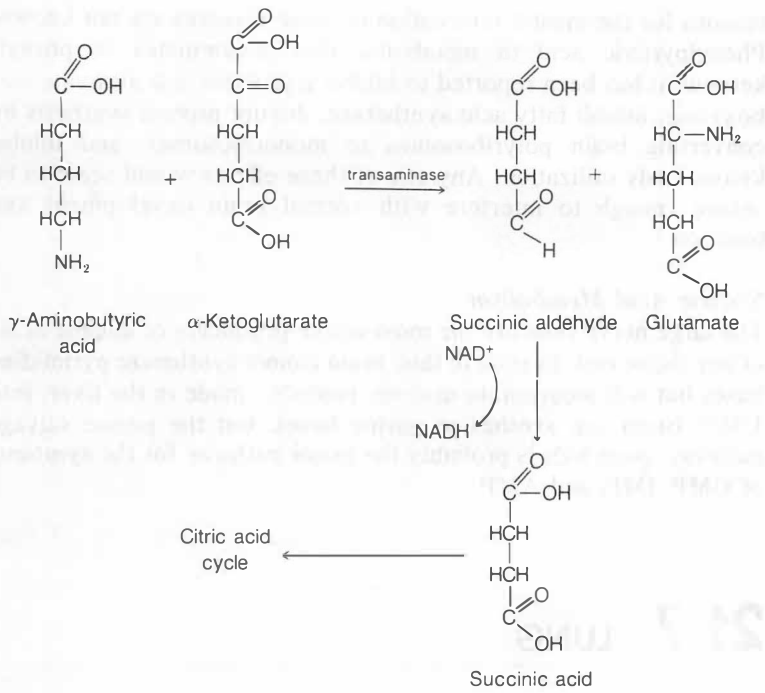


Figure 21.22
The γ-aminobutyric acid (GABA) shunt.

Glutamate and asparate both stimulate cAMP formation from ATP in brain. Two cAMP-dependent protein kinases are found in brain, which stimulate the phosphorylation of proteins in the synaptic membranes. The phosphorylation of synaptic membrane proteins is involved in the transmission of the nervous impulse across the synaptic junction. Biogenic amines (histamine and catecholamine) and a variety of stimuli known to cause depolarization of neural cell membrane also produce marked increases in the levels of cAMP in brain. The elevation of cAMP elicited by glutamate and aspartate is a unique characteristic of the brain cAMP system, in that all other amino acids are ineffective, and brain is the only tissue that responds to the two amino acids. Therefore, glutamate (and/or aspartate) may well be mediator(s) of the phosphorylation of synaptic membrane proteins and thus the transmission of the nervous impulse across the synaptic membrane.

A variety of inborn errors of metabolism have been described in which severe mental retardation occurs involving the amino acids arginine, phenylalanine, tyrosine, histidine, valine, and so on. The

reasons for the mental retardation in these diseases are not known. Phenylpyruvic acid (a metabolite that accumulates in phenylketonuria) has been reported to inhibit glycolysis and pyruvate carboxylase, inhibit fatty acid synthetase, disrupt protein synthesis by converting brain polyribosomes to monoribosomes, and inhibit ketone body utilization. Any one of these effects would seem to be severe enough to interfere with normal brain development and function.

Nucleic Acid Metabolism

The large nerve cells are the most active producers of nucleic acids of any tissue cell. In spite of this, brain cannot synthesize pyrimidine bases but will incorporate uridine, probably made in the liver, into UMP. Brain can synthesize purine bases, but the purine salvage pathway (page 645) is probably the major pathway for the synthesis of GMP, IMP, and AMP.

21.7 LUNG

Glucose is readily metabolized by the lung for energy and for the synthesis of glycogen and lipids. Lung contains a glucose transport system that is stimulated by insulin, leading to increased glucose utilization and lactate production. Fatty acids and ketone bodies are utilized to a very limited extent by isolated lung mitochondria, lung slices, or by perfused lung. The production of lactate accounts for about 60% of the glucose metabolized by lung, and thus this tissue is a major source of blood lactate. The low oxidative capacity of lung is logical in view of its principle role in supplying oxygen to the body. About 10 and 4% of the glucose utilized by lung is by the citric acid cycle and the hexose monophosphate shunt, respectively.

The dependence of lung on insulin for glucose utilization would implicate the necessity for an alternate source of energy in fasting and diabetes. Lung lipoprotein lipase activities and the utilization of fatty acids by lung is increased in fasting. The increase in the oxidative metabolism of lung in fasting and diabetes would decrease the availability of O_2 for the rest of the body. In fact, if the energy for lung metabolism were to arise from the oxidative metabolism of fatty acids, the oxygen consumption of lung would amount to 10% of the total oxygen consumption of a human, an increase of two- to fivefold over normal.

The role of the hexose monophosphate shunt in lung is to furnish

NADPH to be used for the synthesis of fatty acids in surfactant and for the protection against injury to the lung tissue by oxidizing agents, for example, O_2, ozone, and oxygen superoxide (O_2^-).

The Surfactant System of Lung

The surface tension of alveolar fluid is lower than most biological fluids and the integrity of the alveolar space is maintained by the low surface tension of its fluid. A surfactant present in the alveolar fluid is responsible for the low surface tension and the absence or deficiency of this material is the cause of a variety of respiratory diseases. Surfactant is 80–90% lipid, of which phospholipids comprise ~80% of the total lipids. The lipid composition depends on the species of mammal; canine surfactant is shown in Table 21.6. The contributions of the dihydroxyacetone phosphate and the α-glycerol phosphate pathways are about equal for the synthesis of phosphatidylcholine and phosphatidylglycerol. The fatty acids of the surfactant are synthesized in the lung cells. The synthesis of phosphatidylcholine appears to be under the control of glucocorticoids (and perhaps thyroxine) in fetal lung, and this control involves hormone receptor systems. In addition, a protein phosphorylation mechanism, involving cAMP, is a factor in the control of phosphatidylcholine synthesis. Glycoproteins have been isolated from the trachea and bronchi of lung tissue (see Clin. Corrs. 21.4 and 21.5).

Table 21.6 Composition of Canine Surfactant

		Percent of Total
Lipid		91
Saturated phosphatidylcholine	45	
Unsaturated phosphatidylcholine	27	
Phosphatidylglycerol	5	
Phosphatidylethanolamine	2	
Triglyceride	3	
Cholesterol	7	
Other lipid	2	
Protein		7
Carbohydrate		2

SOURCE: From R. J. King, Utilization of alveolar epithelial Type II cells for the study of pulmonary surfactant. *Fed. Proc.*, 38(12):2637–2643, 1979.

CLIN. CORR. 21.4 NEONATAL HYALINE MEMBRANE DISEASE (HMD)

Neonatal collapse of the lung alveoli is responsible for most cases of the respiratory distress syndrome of infants. This disease occurs most frequently in premature infants and in infants of diabetic mothers. Approximately 30% of infants born of women with severe diabetes die of this disease. The number of infant deaths with HMD is estimated to be 25,000 a year in the United States. The cause of HMD is a deficiency in the amount of surfactant in the alveoli of the lungs. The consequences of alveoli collapse are reduced functional residual capacity, lungs become stiff, hypoxia, and respiratory acidosis. Factors predisposing to the development of HMD are prematurity, perinatal asphyxia, diabetic mothers, caesarean section, and genetics.

CLIN. CORR. 21.5 ALVEOLAR PROTEINOSIS (AP)

Alveolar proteinosis may be a response of the lung to a variety of noxious agents, for example, dust. The epithelial cells of the alveoli secrete a proteinaceous material, which fills the alveoli and some of the respiratory bronchioles. The amorphous material consists of three glycoproteins, one of which has been characterized. The excess glycoprotein in the alveoli and bronchioles produces a reduction in vital capacity and the pulmonary compliance, and a reduction in O_2 uptake.

21.8 CONNECTIVE TISSUE

Connective tissue consists of cells and fibers (collagen, elastin, or reticular) imbedded in an amorphous, gelatinous ground substance, which contains the tissue fluids and diffusable metabolites. Examples of connective tissue are the tendons, cartilage, basement membranes, and the intercellular matrix surrounding all cells. The ground substance contains relatively large proportion of proteins, covalently bound to carbohydrates, which give the material a gellike character. The character of the tissue depends on the proportion and the arrangement of the fibers in the ground substance. For example, the intracellular matrix is mainly ground substance with a few fibers, whereas tendon is mainly fibers with minimal ground substance. The structures of the proteoglycans or glycosaminoglycans are discussed in Chapter 8.

The major fibrous tissue in connective tissue is collagen, and in fact, collagen constitutes ~30% of the total body protein. The pathways of synthesis are described in Chapter 19. The structures of the collagen from the various mammalian species and tissues vary greatly. Three varieties of peptide chains have been designated, based on the amino acid composition or sequence, and are called α_1, α_2, and α_3, respectively. The α_1 chains also differ and are called types I, II, III, and IV, respectively. That is, α_1(I), α_1(II), α_1(III), and α_1(IV) chains are found in collagen. Collagen molecules are found that have three identical chains, that is, $[\alpha_1(I)]_3$, $[\alpha_1(II)]_3$, $[\alpha_1(III)]_3$, and $[\alpha_1(IV)]_3$, and that have two identical and one nonidentical chains, for example, $[\alpha_1(I)]_2\alpha_2$. Examples of different collagens are in Table 21.7. αA and αB collagens are a newly discovered "maintenance type" of collagen found in low amounts in various tissues. The above classifications of collagens are based on the amino acid sequence and composition of the peptide chains in the tropocollagen molecule. The compositions of the collagen molecules also differ in the degree of hydroxylation of proline and lysine, the extent of the glycosylation of hydroxylysine, the number of aldehyde groups on the chain, and the length of the amino terminal, nonhelical peptides. In other words, collagen is a *large* family of molecules.

Basement Membranes

Basement membranes occur as extracellular components of certain tissues: lining of vascular tree capillaries, kidney tubules, lung alveoli, lens capsule of the eye, and kidney glomerulus. The basement membranes contain collagen in combination with a glycoprotein.

Table 21.7 Structurally and Genetically Distinct Collagens

Type	*Tissue Distribution*	*Molecular Form*	*Chemical Characteristics*
I	Bone, tendon, skin, dentin, ligament, fascia, arteries, and uterus	$[\alpha 1(I)]_2\alpha 2$	Hybrid composed of 2 kinds of chains: low in hydroxylysine and glycosylated hydroxylysine
II	Hyaline cartilage	$[\alpha 1(II)]_3$	Relatively high in hydroxylysine and glycosylated hydroxylysine
III	Skin, arteries and uterus	$[\alpha 1(III)]_3$	High in hydroxyproline and low in hydroxylysine; contains interchain disulfide bonds
IV	Basement membranes	$[\alpha 1(IV)]_3$	High in hydroxylysine and glycosylated hydroxylysine; may contain large globular regions
V	Basement membranes and perhaps other tissues	αA and αB	Similar to type IV

SOURCE: From D. J. Prockop, K. I. Kivirikko, L. Tuderman, and N. A. Guzman, The biosynthesis of collagen and its disorders. *New Engl. J. Med.*, 301:13–33, 1979.

The glycoprotein (mol wt ~70,000) contains sialic acid, fucose, galactose, mannose, and hexosamine. The protein consists mostly of low amounts of hydroxyproline and hydroxylysine, and large amounts of acidic amino acids, cystine, and glycine. The collagen portion contains less glycine and more hydroxylysine and hydroxyproline than other collagens, and the hydroxalated amino acids are highly glycosylated. The molecular weight of collagen in basement membrane is ~120,000, which is larger than that of most collagen, which is ~95,000. The collagen and glycoprotein may be linked together by covalent linkages involving disulfide bonds, or the glycoprotein may be an extension of the procollagen molecule.

Turnover of Collagen

The cells of the connective tissue responsible for the synthesis of the fibers and ground substance are the chrondocytes, fibroblasts, and mast cells. Collagenases initiate a physiological and pathological

CLIN. CORR. 21.6 GLYCOSAMINOGLYCANS

At least nine inborn errors of metabolism have been identified in which enzymes that degrade the glycosaminoglycans are deficient and are listed in Table 21.8. The breakdown of the proteoglycans (covalent protein glycosaminoglycan complex) begins in the tissue. The glycosaminoglycans are further broken down by proteases and glycosidases in the tissue cells. The degradation of the glycosaminoglycans is completed in the liver lysosomes by the formation of products that are excreted in the urine, that is, chondroitin and heparitin sulfate, hyaluronic acid, and so on. The genetic diseases of collagen are described in Chapter 19.

breakdown of collagen by extracellular cleavage of the helical portion of native collagen molecules at about three-quarters the distance from the N terminus. Collagenase is released from a variety of tissue cells, for example, skin and bone, which is later activated by extracellular proteolysis. The turnover rate of collagen varies with its location: for example, the turnover rate of collagen in the gingiva is greater than it is in bone, skin, and cartilage, with tendon having the lowest turnover rate (see Clin. Corr. 21.6).

21.9 EYE

Lens

The lens is a colorless, transparent, vascular tissue, composed entirely of epithelial cells. The anterior surface consists of a layer of intact cells, whereas the rest of the lens consists of cells that progressively lose their mitochondria and nuclei—the loss of which is a factor that contributes to the transparency of the tissue. Mammalian lens consists of ~35% protein and four different proteins have been identified, called α, β, and γ crystallins, and albuminoid protein. The source of the lens nutrients and the route whereby its waste products are removed is a slowly circulating aqueous humor that bathes its front surface. The lens uses ATP for maintenance, growth and repair, protein synthesis, and transport of cations (via a Na^+,K^+-ATPase) and amino acids. If ATP production by the lens is disrupted, the lens quickly accumulates Na^+ and H_2O and becomes opaque. Most of the energy for ATP synthesis in the lens arises by glycolysis, and the percent glucose metabolized by the glycolytic

Table 21.8 Genetic Diseases Involving Glycosaminoglycans

Missing Enzyme	*Deficiency Disease*
Iduronate sulfatase	Hunter syndrome
α-L-Iduronidase	Hunter and Scheie syndromes
N-Acetylgalactosamine sulfatase	Maroteaux–Lamy syndrome
N-Acetyl β-galactosaminidase	Unknown
β-Glucuronidase	A mucopolysaccharidosis
Heparan *N*-sulfatase	Sanfillipo syndrome, type A
α-Glucosaminidase	Unknown
N-Acetylglucosamine sulfatase	Morquio syndrome
N-Acetyl α-glucosaminidase	Sanfillipo syndrome, type B

and hexose monophosphate shunt is ~85 and 10%, respectively. Approximately 3% of the glucose is completely oxidized to CO_2 and H_2O by the citric acid cycle. The control of glycolysis by the lens is critical in that excess lactate damages the lens. The control step in lens glycolysis is hexokinase, which is inhibited by glucose 6-phosphate. A decrease in the amount of ATP in lens activates phosphofructokinase, which in turn produces a decrease in the amount of glucose 6-phosphate. The enzyme hexokinase is stimulated in the presence of low amounts of glucose 6-phosphate. An increase in ATP in lens has the opposite, or an inhibitory effect, on hexokinase. The NADH from glycolysis in the lens is used mainly for the conversion of pyruvate to lactate, with small amounts of reducing equivalents transferred to the few mitochondria present in lens by the α-glycerophosphate shuttle (page 296).

The only disease of lens is the formation of cataracts, an opacity of the lens. Cataracts can be produced experimentally in animals fed large amounts of galactose or sorbitol and are found in humans with hyper- and hypoglycemia and with deficiencies in galactokinase. Low glucose levels in the blood result in the destruction of lens hexokinase, thereby decreasing the glycolytic rate, and high amounts of glucose in the lens activate the sorbitol pathway leading to fructose formation. The sorbitol pathway is shown in Figure 21.23.

Retina

The retina contains the enzymes of glycolysis, citric acid cycle, and hexose monophosphate shunt. Retina has one of the most active glycolytic systems and the highest rate of oxygen consumption of any tissue. The retina contains large amounts of lactate and lactate dehydrogenase, which means that the rate of pyruvate production by glycolysis exceeds its rate of oxidative metabolism even under conditions of adequate oxygen supplies. Low activities of the α-glycerophosphate shuttle enzymes (page 296) and the large amounts of lactate in retina are interpreted to indicate the NADH produced in glycolysis is used for the reduction of excess pyruvate to lactate. Lactate dehydrogenase of retina utilizes NADP as well as NAD; thus the production of NADPH from lactate is a factor in controlling the rate of glucose metabolism by the hexose monophosphate shunt. The production of NADPH by the hexose monophosphate shunt is a critical reaction for retina, in view of its utilization in the *Rhodopsin cycle*.

Role of Rhodopsin in Vision

The retina contains two types of receptor cells: cones, which are specialized for color and detail vision in bright light, and rods, which

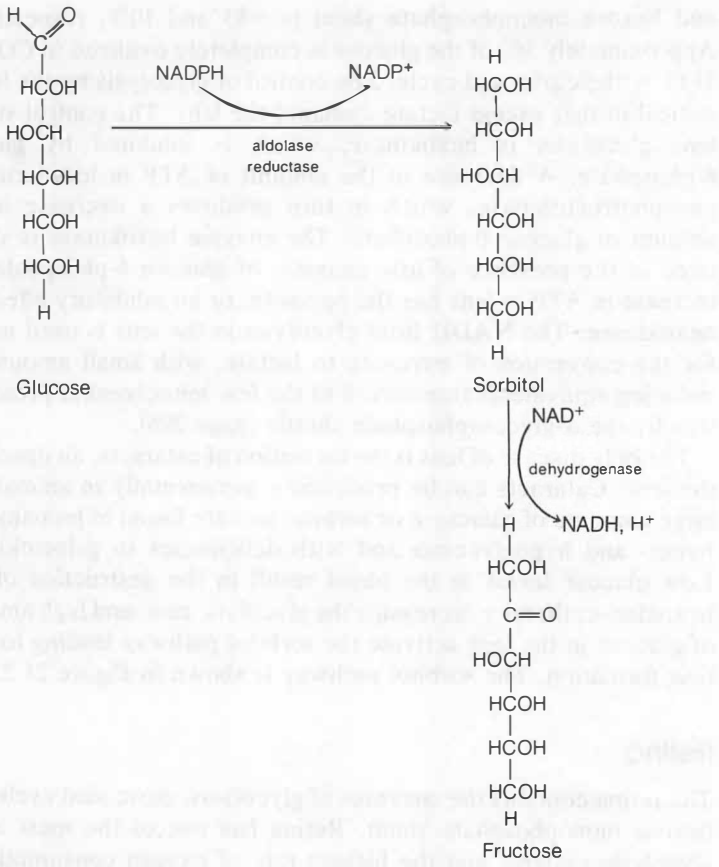


Figure 21.23
The sorbitol pathway.

are specialized for vision in dim light. Light waves striking these receptors produce chemical changes that give rise to nervous impulses that are transmitted to the brain. Vision in the rod cells is dependent upon a photosensitive pigment called rhodopsin or *visual purple*. Rhodopsin is a complex of the protein opsin and the 11-*cis* isomer of the aldehyde form of vitamin A, called retinal. Opsin is a hydrophobic protein (mol wt ~28,000). When light strikes the rhodopsin in the retina, the 11-*cis*-retinal is isomerized to the all-*trans* form yielding an all-*trans*-chromoprotein complex called prelumirhodopsin (Figure 21.24). This is the only step in the rhodopsin cycle that is affected by light and is the step that is believed to be associated with the nervous impulse. The prelumirhodopsin under-

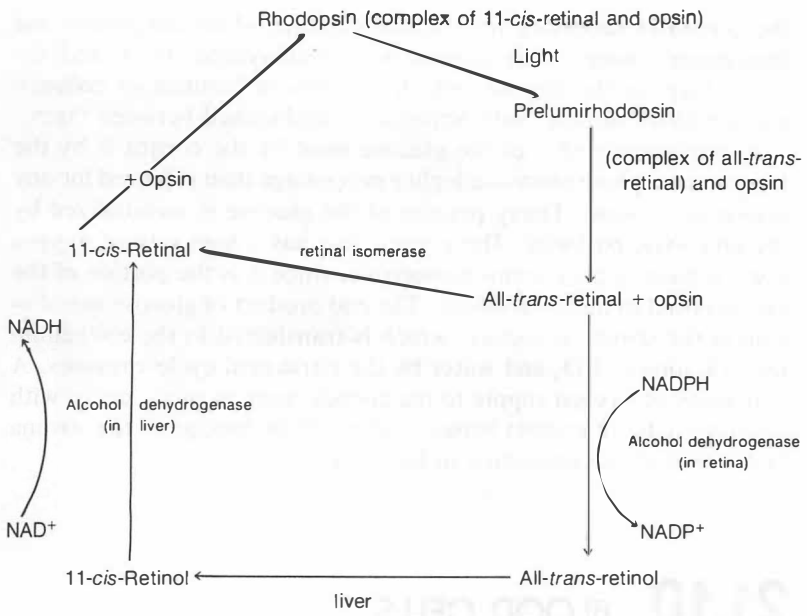


Figure 21.24
The rhodopsin cycle.

goes a series of reactions, resulting in the cleavage of the complex to all-*trans*-retinal and the protein opsin.

Rhodopsin must be reconstituted for continued vision in dim light. The all-*trans* isomer of retinal is converted to 11-*cis*-retinal by the action of retinal isomerase in the retina, and the 11-*cis*-retinal combines with opsin to form rhodopsin. The all-*trans*-retinal has an alternate fate; it is reduced to all-*trans*-retinol by alcohol dehydrogenase and NADPH in the retina. The all-*trans*-retinol is transported to the liver as its fatty acid esters, and both the all-*trans*-retinol and its esters are isomerized to 11-*cis*-retinol and oxidized to 11-*cis*-retinal to be used for the formation of rhodopsin. The isomerization of all-*trans*-retinol to 11-*cis*-retinol occurs in the liver.

Cornea

The cornea, like the lens, is an avascular organ and mainly derives its nutrients from the aqueous humor. The cornea also derives nourishment from limbus, tears, and from vessels in the lids. The single cell layer of the cornea endothelium contains the transport systems for nutrients and water. The constant removal of water from

the cornea is necessary for the maintenance of an anhydrous and transparent cornea. The cornea is a multilayered tissue and the major layer is the stroma, which consists of laminae of collagen parallel to its surface, with fibroblasts sandwiched between them.

Approximately 65% of the glucose used by the cornea is by the hexose monophate shunt—a higher percentage than reported for any mammalian tissue. Thirty percent of the glucose is metabolized by the glycolytic pathway. The cornea also has a high rate of oxygen consumption, which seems reasonable, since it is the portion of the eye exposed to the atmosphere. The end product of glucose metabolism in the stroma is lactate, which is transferred to the epithelium for oxidation to CO_2 and water by the citric acid cycle enzymes. A deficiency of oxygen supply to the cornea, such as might occur with improperly fitted contact lenses, will result in damage to the stroma because of an accumulation of lactic acid.

21.10 BLOOD CELLS

Reticulocyte

The immature red blood cell, called the reticulocyte, contains the subcellular organelles of all mammalian cells except a nucleus. The reticulocyte has the enzymes involved in glycolysis; hexose monophosphate shunt, citric acid cycle; oxidative phosphorylation; synthesis of hemoglobin, cholesterol, phospholipids, triglycerides, and purine nucleotides. Reticulocytes have higher rates of glucose consumption than erythrocytes and when cellular respiration is inhibited, both the glucose uptake and lactate production of the reticulocyte are greatly increased. The Pasteur effect in reticulocytes is greater than that reported for any other mammalian tissue. The increase in glycolytic rate of reticulocytes in a lack of oxygen is attributed to an activation of phosphofructokinase and pyruvate kinase. Reticulocytes also have the enzyme glycerol kinase and can utilize glycerol for the production of energy.

Erythrocyte

The mitochondria, Golgi apparatus, and RNA are extruded from the reticulocyte in the process of maturation to form the erythrocyte, and therefore erythrocytes are not living cells in the strict sense. Erythrocytes cannot synthesize protein, glycogen, lipids (from glucose), and are incapable of oxidative metabolism. They utilize glu-

cose by glycolysis and the hexose monophosphate shunt exclusively and have the ability to synthesize nucleotides from preformed purines. The erythrocytes survive and carry out their O_2 transport function for ~120 days after which time they are broken down in the body.

Erythrocytes contain ~35% solid material, which is almost all hemoglobin (32%). The remaining solids are proteins and lipids which form the stromal network for the support of the hemoglobin. The membrane is freely permeable to water, urea, creatinine, HCO_3^-, Cl^-, OH^-, and H^+. The transport of Na^+, K^+, Ca^{2+}, and perhaps P_i across the erythrocyte membrane requires energy. The phosphate content of the RBC is 50–100 fold greater than in plasma, most of which is organic, that is, hexose phosphate, triose phosphates, and ATP.

The metabolic fuel of the erythrocyte is glucose which is utilized by glycolysis. The energy of ATP is necessary to preserve the integrity of the erythrocyte and is also used for the membrane transport of ions. The NADH from glycolysis is used for the conversion of methemoglobin to hemoglobin and for the conversion of pyruvate to lactate. Methemoglobin is slowly but continually formed from hemoglobin in the body and is not capable of O_2 transport. Hexokinase, phosphofructokinase, pyruvate kinase, glyceraldehyde 3-phosphate dehydrogenase, and phosphoglycerate kinase are implicated in the control of glycolysis in the erythrocyte. A number of hemolytic anemias have been identified due to deficiencies of these enzymes (see Clin. Corr. 21.7). A major factor in the activation of glycolysis in the erythrocyte is inorganic phosphate. Approximately 10% of the glucose metabolized by the erythrocyte is by the hexose monophosphate shunt. The ratio of the oxidized to reduced form of glutathione, present in large amounts in the erythrocyte, the amount of H_2O_2 in the RBC, and a breakdown product of NADP, 2′-phosphoadenosine diphosphate ribose, are implicated in the control of the shunt. The glutathione and H_2O_2 may well control the rate of shunt metabolism by altering the ratio of the reduced to oxidized form of NADP.

A carrier protein is involved in the transport of glucose in erythrocytes and constitutes a major component of the erythrocyte membrane. The movement of glucose does not require energy or insulin. The number of sugar carrier sites per red cell has been caculated to be 6×10^4. An interchange of cholesterol and phospholipids between plasma and the erythrocyte membrane occurs, which is a nonenzymatic process that is a factor in maintaining the integrity of the membrane.

A cAMP dependent protein kinase is found in human erythrocyte membrane. The function of this enzyme is unknown, since cAMP

CLIN CORR 21.7 HEMOLYTIC ANEMIA

Pyruvate kinase deficiency hemolytic anemia is the result of a deficiency of this enzyme in the crythrocyte and is an inherited disease. A second commonly found hemolytic anemia is the result of a deficiency of glucose 6-phosphate dehydrogenase in the RBC. Other hereditary hemolytic anemias are the result of deficiencies in the RBC of each of the following enzymes: hexokinase, glucosephosphate isomerase, phosphofructokinase, triosephosphate isomerase, 2,3-diphosphoglyceromutase, phosphoglycerate kinase, 6-phosphogluconate dehydrogenase, glutathione reductase, glutathione peroxidase, and glutathione synthetase.

does not greatly stimulate protein kinase activity in the erythrocyte, and the erythrocyte has low levels of cAMP.

The modulating effect of 2,3-diphosphoglyceric acid on the combination of O_2 with hemoglobin is discussed on page 1093.

Leukocytes

The polymorphonuclear leukocytes possess the enzymes of the hexose monophosphate shunt, glycolysis, citric acid cycle, and respiratory enzymes. Only a small amount of glucose is oxidized via citric acid cycle and respiration. The primary purpose of the leukocyte is phagocytosis, which is the engulfing and destruction of particulate matter. Phagocytosis is an energy-requiring process associated with an increase in the rate of glycolysis and a great increase in the metabolism of glucose by the hexose monophosphate shunt. The role of the hexose monophosphate shunt is to produce H_2O_2 (from superoxide, O_2^-), which is used in the phagocytotic process. Large amounts of glycogen are found in leukocytes, which is the energy source for phagocytosis in the absence of exogenous glucose. If glucose is present, glycogen breakdown does not occur in phagocytosis.

Platelets

Platelets consist of 50% protein and 15% lipid and contain a contractile protein (similar to actomyosin of muscle) and ATP. The actomyosin and ATP are involved in the clot retraction mediated by the platelets. Platelets contain substances required in blood clotting and enzymes characteristic of lysosomes and mitochondria.

The major energy source for platelets arises from the glycolysis of glucose. Under aerobic conditions, glycolysis accounts for one-half the ATP formed. Platelets increase the rate of glucose utilization and lactate production in a deficiency of oxygen. Approximately 20% of the glucose metabolized by platelets is by the hexose monophosphate shunt. The aggregation of platelets in blood clotting is an energy-requiring process, which is derived from its breakdown of glycogen and glycolysis.

21.11 SKIN

The skin consists of the outer epidermis and its supporting inner connective tissue, the dermis. The epidermis gives rise to hair, nails,

and sweat glands, and all are considered to be continuous with the epidermis. The dermis contains the blood vessels, nerves, and lymphatics of the skin, and encloses the hair follicles, sebaceous glands, and sweat glands. Human skin accounts for about 10% of the body weight.

Approximately 80% of the glucose is metabolized by glycolysis, and about one-half of the remaining glucose is metabolized by the hexose monophosphate shunt and the other half is converted to glycogen. Glucose utilization by the epidermis is an insulin-dependent process. The source of energy for skin is glycolysis, and skin contributes large amounts of lactate to the blood. The concentration of lactate in skin is about 10-fold greater than in plasma. The control of glycolysis in skin is considered to be by the enzyme phosphofructokinase. The epidermis is capable of synthesizing DNA, RNA, and protein—which is principally keratin—a highly insoluble protein. A surface film of lipid containing large amounts of cholesterol and squalene is synthesized in the epidermal cells. The conversion of 7-dehydrocholesterol to activated 7-dehydrocholesterol occurs in the epidermis.

Pigment-forming cells called melanocytes occur in the epidermis and are responsible for the formation and secretion of melanin. The first step in melanin synthesis is the hydroxylation of tyrosine to form 3,4-dihydroxyphenylalanine (dopa) which then undergoes a series of reactions to form compounds that polymerize to form melanin (see Clin. Corr. 21.8).

CLIN. CORR. 21.8 ALBINISM

Albinism is the clinical manifestation of an inheritable metabolic defect in the pigment cell (melanocyte) system of the eye and skin. Patients with albinism of the skin lack the ability to convert tyrosine to melanin, and their skin is not pigmented.

BIBLIOGRAPHY

General

Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. *The metabolic basis of inherited disease*, 3rd ed. New York: McGraw-Hill, 1972.

Skeletal Muscle

Ebashi, S. Excitation-contraction coupling. *Annu. Rev. Physiol.* 38:293, 1976.

Huxley, H. E. The mechanism of muscular contraction. *Sci. Am.* 213(6):18, 1965.

Huxley, H. E. Regulation of muscle function by tropomyosin-troponin, in Y. Hatefi and L. Djavadi-Ohaniance (eds.), *The structural basis of membrane function*. New York: Academic, 1976, p. 313.

MacLennan, D. H. Resolution of the calcium transport system of the sacroplasmic reticulum. *Can. J. Biochem.* 53:251, 1972.

Taylor, E. W. Chemistry of muscle contraction. *Annu. Rev. Biochem.* 41:577, 1972.

Adipose Tissue

Kinsell, L. W. (ed.). Adipose tissue as an organ. Proceedings of the Duel Conference of Lipids. Springfield, Ill.: C. C. Thomas, 1962.

Liver

Brodie, B. B., and Gillette, J. R. (eds.). *Concepts in biochemical pharmacology,* part 2: *Handbook of experimental pharmacology.* New York: Springer-Verlag, 1971.

Kidney

Anders, M. W. Metabolism of drugs by kidney. *Kidney Int.* 18:636, 1980.

Cohen, J. J., and Barac-Nieto, M. Renal metabolism of substrates in relation to renal function. *Handbook of physiology,* Section 8, 909, 1973.

Nervous Tissue

Albers, R. W., Agranoff, B. W., Katzman, R., and Siegel, G. J. (eds.). *Basic neurochemistry.* Boston: Little, Brown, 1972.

McLlwain, H., and Bachelard, H. S. *Biochemistry and the nervous system,* 4th ed. London: Churchill-Livingston, 1973.

Morell, P., and Norton, W. T. Myelin. *Sci. Am.* 242:88, 1980.

Lung

Cohen, A. B. Lung metabolism: cells. *Fed. Proc.* 38:2635, 1979.

Heinemann, H. O. Metabolism of lung. *Fed. Proc.* 32:1955, 1973.

Eye

Rathbun, W. B. Biochemistry of the lens and cataractogenesis: current concepts, in R. L. Peiffer (ed.), *Veterinary clinics of North America: small animal practice.* Philadelphia: W. B. Saunders, 110, 377, 1980.

Blood Cells

Harris, J. W., and Kellermeyer, R. W. *The red cell.* Cambridge, Mass.: The Commonwealth Fund, Harvard University Press, 1972.

22

Iron and Heme Metabolism

WILLIAM M. AWAD, JR.
MARILYN S. WELLS

22.1 IRON METABOLISM: OVERVIEW

Iron is closely involved in the metabolism of oxygen; the properties of this metal permit the transportation and participation of oxygen in a variety of biochemical processes. The common oxidation states are either ferrous (Fe^{2+}) or ferric (Fe^{3+}); higher oxidation levels occur as short-lived intermediates in certain redox processes. Iron has an affinity for electronegative atoms such as oxygen, nitrogen, and sulfur. These latter atoms provide the electrons that form bonds with iron, and these can be of very high affinity when favorably oriented on macromolecules. In forming complexes no bonding electrons are derived from iron. There is an added complexity to the structure of iron: the nonbonding electrons in the outer shell of the metal (the incompletely filled 3*d* orbitals) can exist in two states. Where bonding interactions with iron are weak, the outer nonbonding electrons will avoid pairing and distribute throughout the 3*d* orbitals. Where bonding electrons interact strongly with iron, however, there will be pairing of the outer nonbonding electrons, favoring lower energy 3*d* orbitals. These two different distributions for each oxidation state of iron can be determined by electron spin resonance measurements. Dispersion of 3*d* electrons to all orbitals leads to the high-spin state, whereas restriction of 3*d* electrons to lower energy orbitals leads, because of electron pairing, to a low-spin state. Some iron–protein complexes reveal changes in spin state without changes in oxidation during chemical events (e.g., binding and release of oxygen by hemoglobin).

At neutral and alkaline pH ranges, the redox potential for iron in aqueous solutions favors the ferric state; at acid pH values, the equilibrium favors the ferrous state. In the ferric state iron slowly forms large polynuclear complexes with hydroxide ion, water, and other anions that may be present. These complexes can become so large as to exceed their solubility products, leading to their aggregation and precipitation with pathological consequences.

Iron can bind to and influence the structure and function of a variety of macromolecules, with deleterious results to the organism. To protect against such reactions several iron-binding proteins function specifically to store and transport iron. These proteins have both a very high affinity for the metal and, in the normal physiological state, also have incompletely filled iron-binding sites. The interaction of iron with its ligands has been well characterized in some proteins (e.g., hemoglobin and myoglobin), whereas for others (e.g., transferrin) it is presently in the process of being defined. The major

area of ignorance in the biochemistry of iron lies in the almost complete lack of information on the in vivo transfer of iron from one macromolecule to another. Three mechanisms have been proposed to explain the process of iron transfer. Two of these have been supported by excellent model studies but with varying degrees of relevance to the physiological state. The three proposed processes are the following. First, redox changes of iron has been an attractive mechanism because it is supported by selective in vitro studies and because in many cases macromolecules show a very selective affinity for ferric ions, binding ferrous ions poorly. Thus reduction of iron would permit ferrous ions to dissociate and reoxidation would allow the iron to redistribute to appropriate macromolecules. Redox mechanisms have only been defined in a very few settings, some of which will be described below. An alternative hypothesis involves chelation of ferric ions by specific small molecules with high affinities for iron; this mechanism has been supported also by selective in vitro studies. The chelation mechanism suffers from the lack of a demonstrably specific in vivo chelator. A third mechanism, which has been studied less well, is the possibility that conformational transitions are imposed upon iron-containing proteins leading to the loosening of ligands to iron. Because the redox potential strongly favors ferric ion at almost all tissue sites and because Fe^{3+} binds so strongly to liganding groups, the probability is that there are cooperating mechanisms regulating the transfer of iron.

22.2 IRON-CONTAINING PROTEINS

Iron binds to proteins either by incorporation into a protoporphyrin IX ring (see below) or by interaction with other protein ligands. Ferrous- and ferric-protoporphyrin IX complexes are designated, respectively, heme and hematin. Heme-containing proteins include those that transport (e.g., hemoglobin) and store (e.g., myoglobin) oxygen; those involved in the electron transport system (e.g., cytochromes); and certain enzymes which contain heme as part of their prosthetic groups (e.g., tryptophan pyrrolase). Discussions on structure–function relationships of heme proteins are presented in Chapters 6 and 23.

Nonheme proteins include transferrin, ferritin, and a variety of redox enzymes that contain iron at the active site, and also iron-sulfur proteins. A significant body of information has been acquired

which relates to the structure–function relationships of some of these molecules.

Transferrin

The major protein in serum involved in the transport of iron is transferrin, a β-1-glycoprotein synthesized in the liver, consisting of a single polypeptide chain of 78,000 daltons with two iron-binding sites. Sequence studies indicate that the protein is a product of gene duplication derived from a putative ancestral gene coding for a protein binding only one atom of iron. Though several other metals bind to transferrin, the highest affinity is for ferric ion. Transferrin does not bind ferrous ion. The binding of each ferric ion is absolutely dependent upon the coordinate binding of an anion, which in the physiological state is carbonate as indicated below.

$$\text{Transferrin} + Fe^{3+} + CO_3^{2-} \longrightarrow \text{transferrin} \cdot Fe^{3+} \cdot CO_3^{2-}$$

$$\text{Transferrin} \cdot Fe^{3+} \cdot CO_3^{2-} + Fe^{3+} + CO_3^{2-} \longrightarrow \text{transferrin} \cdot 2(Fe^{3+} \cdot CO_3^{2-})$$

In experimental settings, other organic polyanions can substitute for carbonate. Estimates of the association constants for the binding of ferric ion to transferrins from different species range from 10^{24} to 10^{31} M^{-1}, indicating for practical purposes that wherever there is excess transferrin free ferric ions will not be found. There is no evidence for cooperativity in the binding of iron at the two sites. In the normal physiological state, approximately one-ninth of all transferrin molecules are saturated with iron at both sites; four-ninths of transferrin molecules have iron at either site; and four-ninths of circulating transferrin are free of iron. The two iron-binding sites, though homologous, are not completely identical; they show some differences in sequences and also some differences in affinities for other metals (especially the lanthanides). The two iron sites have different physiological roles with independent specificities for the delivery of iron to different peripheral tissues. This matter has not been completely resolved. Removal of terminal sialyl residues from the oligosaccharide attached to transferrin does *not* lead to rapid uptake of the modified protein by hepatocytes. This is the major exception to the rule that exposure of penultimate galactosyl residues following the removal of terminal sialyl residues leads to a glycoprotein's uptake by the mammalian liver.

Lactoferrin

Milk contains iron that is almost exclusively bound to a protein, which is closely homologous to transferrin with two sites binding the

metal. The iron content of lactoferrin varies, but the protein is never saturated. Surprisingly, the role of lactoferrin in facilitating the transfer of iron to intestinal receptor sites in the infant has not been carefully examined. Rather, major studies on the function of lactoferrin have been directed toward its antimicrobial effect, protecting the newborn from gastrointestinal infections. Microorganisms require iron for replication and function. The presence of incompletely saturated lactoferrin results in the rapid binding of any free iron leading to the inhibition of microbial growth by preventing a sufficient amount of iron from entering these microorganisms. Other microbes, such as *Escherichia coli*, which release competitive iron chelators, are able to proliferate despite the presence of lactoferrin, since the chelators transfer the iron specifically to the microorganism.

Ferritin

This is the major protein involved in the storage of iron. The protein consists of an outer polypeptide shell 130 Å in diameter with a central ferric-hydroxide-phosphate core, 60 Å across. The apoprotein has a molecular weight of 440,000 and consists of 24 subunits of 18,500 daltons each. The ratio of iron to polypeptide is not constant, since the protein has the ability to gain and release iron according to physiological needs. Usually the average iron content per molecule is less than 3,000 atoms; there is a capacity of ~4,500 iron atoms per ferritin sphere. The outer protein shell has channels to the core permitting accumulation, and release of iron. When iron is in excess, the storage capacity of newly synthesized apoferritin may be exceeded. This leads to iron deposition adjacent to ferritin spheres. Histologically such amorphous iron deposition is called hemosiderin.

Ferritin facilitates the oxidation of ferrous ions to the ferric state. Ferritins derived from different tissues of the same species demonstrate differences in electrophoretic mobility in a fashion analogous to the differences noted with isoenzymes. In some tissues ferritin spheres form latticelike arrays, which are easily identifiable by electron microscopy. The synthesis of ferritin is inhibited by apoferritin molecules. Release of the apoferritin molecules from the polyribosome occurs only after the incorporation of iron. Thus the association of apoferritin with polyribosomes leads to inhibition of synthesis of the protein when it is not needed (see Figure 22.1).

Other Nonheme Iron-Containing Proteins

Many iron-containing proteins are involved in enzymatic processes, most of which are related to oxidation mechanisms. The structural

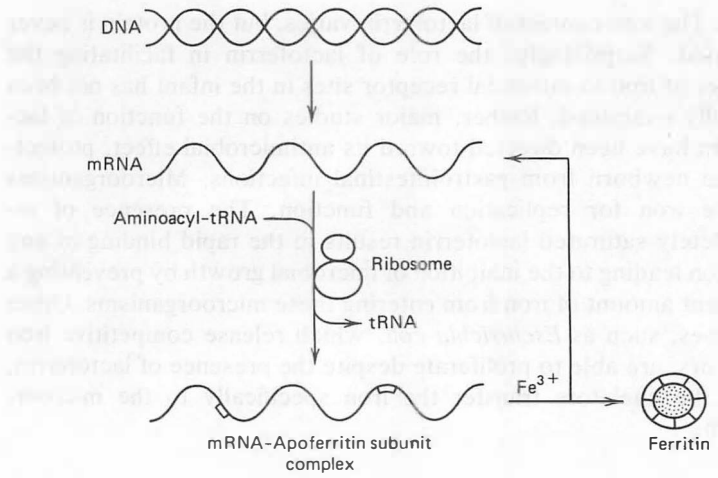


Figure 22.1
Translational control of apoferritin synthesis.

features of the ligands binding the iron are not well-known, except for a few components involved in mitochondrial electron transport. These latter proteins are characterized by iron being bonded only to sulfur atoms. Three major types of such proteins are known (see Figure 22.2). The smallest type (e.g., nebredoxin) found only in

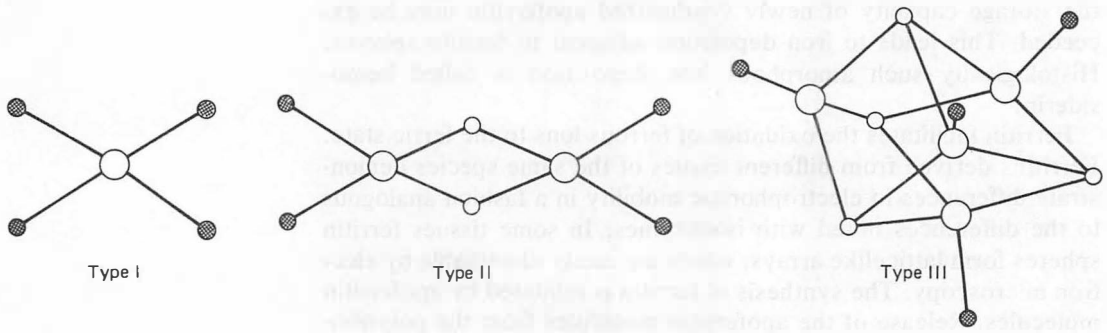


Figure 22.2
Structure of the ferredoxins.
The large open circles represent the iron atoms; the small open circles represent the inorganic sulfur atoms; and the small stippled circles represent the cysteinyl sulfur atoms derived from the polypeptide chain.

microorganisms, consists of a small polypeptide chain with a molecular weight of about 6,000, containing one iron atom bound to four cysteine residues. A second type consists of ferredoxins found in both plants and animal tissues where two iron atoms are found, each liganding to two separate cysteine residues and sharing two sulfide anions. The most complicated of the iron-sulfur proteins are the bacterial ferredoxins, which contain four atoms of iron, each of which is linked to single separate cysteine residues but also shares three sulfide anions with neighboring iron molecules to form a cubelike structure. In some anaerobic bacteria a family of ferredoxins may contain two of the third type of iron-sulfur groups per macromolecule. The redox potential afforded by these different ferredoxins varies widely and is in part dependent upon the environment of the surrounding polypeptide chain that envelops these iron-sulfur groups. In nebredoxin the iron undergoes ferric-ferrous conversion during electron transport. With the plant and animal ferredoxins (type II iron-sulfur proteins) both irons are in the ferric form in the oxidized state; upon reduction only one iron goes to the ferrous state. In the bacterial ferredoxin (type III iron-sulfur protein) the oxidized state can be either $2Fe^{3+} \cdot 2Fe^{2+}$ or $3Fe^{3+} \cdot Fe^{2+}$, with corresponding reduced forms of $Fe^{3+} \cdot 3Fe^{2+}$ or $2Fe^{3+} \cdot 2Fe^{2+}$.

22.3 INTESTINAL ABSORPTION OF IRON

The high affinity of iron for both specific and nonspecific macromolecules leads to the absence of significant formation of free iron salts, and thus this metal is not lost via usual excretory routes. Rather, excretion of iron occurs only through the normal sloughing of tissues that are not reutilized (epidermis and gastrointestinal mucosal cells). In the healthy adult male the loss is ~1 mg/day. In premenopausal women, the normal physiological events of menses and parturition substantially augment iron loss. A wide variation of such loss exists, depending upon the amounts of menstrual flow and the multiplicity of births. In the extremes of the latter settings, a premenopausal woman may require an amount of iron that is four to five times that needed in an adult male for prolonged periods of time. The postmenopausal woman who is not iron-deficient has an iron requirement similar to that of the adult male. Children and patients with blood loss have naturally increased iron requirements.

Cooking of food facilitates the breakdown of ligands attached to

iron, increasing the availability of the metal in the gut. The low pH of stomach contents permits the reduction of ferric ion to the ferrous state, facilitating dissociation from ligands. The latter requires the presence of an accompanying reductant, which is usually achieved by adding ascorbate to the diet. The absence of a normally functioning stomach reduces substantially the amount of iron that is absorbed. Some iron-containing compounds bind the metal so tightly that it is not available for assimilation. Thus, contrary to popular belief, spinach is a very poor source of iron because the iron content is bound almost exclusively to phytate (hexaphos-

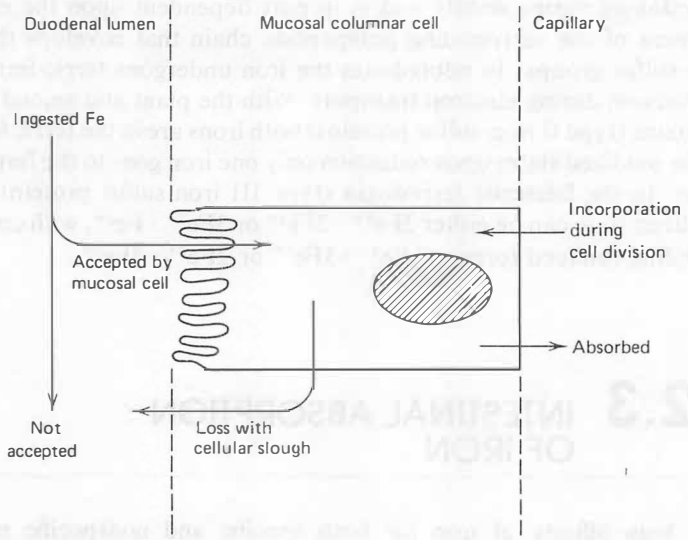


Figure 22.3
Intestinal mucosal regulation of iron absorption.
The flux of iron in the duodenal cell is indicated. A fraction of the iron that is potentially acceptable is transferred from the intestinal lumen into the epithelial cell. A large portion of ingested iron is not absorbed, in part because it is not presented in a readily acceptable form. Some iron is retained within the cell, bound by apoferritin to form ferritin. This iron is sloughed into the intestinal lumen with the normal turnover of the cell. A portion of the iron within the mucosal cell is absorbed and transferred to the capillary bed to be incorporated into transferrin. During cell division, which occurs at the bases of the intestinal crypts, iron is incorporated for cellular requirements. These fluxes change dramatically in iron-depleted or iron-excess states.

phoinositol), which is resistant to the chemical actions of the gastrointestinal tract. Specific protein cofactors derived from the stomach or pancreas have been suggested as being facilitators of iron absorption in the small intestine.

The major site of absorption of iron is in the small intestine with the largest amount being absorbed in the duodenum and a gradient of lesser absorption occurring in the more distal portions of the small intestine. The metal enters the mucosal cell either as the free ion or as heme; in the latter case the metal is split off from the porphyrin ring in the mucosal cytoplasm. The large amount of bicarbonate secreted by the pancreas neutralizes the acidic material delivered by the stomach and thus favors the oxidation of ferrous ion to the ferric state. The major barrier to the absorption of iron is not at the lumenal surface of the duodenal mucosal cell. Whatever the requirements of the host are, in the face of an adequate delivery of iron to the lumen, a substantial amount of iron will enter the mucosal cell. Regulation of iron transfer occurs between the mucosal cell and the capillary bed (see Figure 22.3). In the normal state, certain processes define the amount of iron that will be transferred. Where there is iron deficiency, the amount of transfer increases; where there is iron overload in the host, the amount transferred is curtailed substantially. One mechanism that has been demonstrated to regulate this transfer of iron across the mucosal–capillary interface is the synthesis of apoferritin by the mucosal cell. Where little iron is required by the host, a large amount of apoferritin is synthesized to trap the iron within the mucosal cell and prevent transfer to the capillary bed. As the cells turn over (within a week), their contents are extruded into the intestinal lumen without absorption occurring. Where there is iron deficiency, virtually no apoferritin is synthesized, so as not to compete against the transfer of iron to the deficient host. There are other as yet undefined positive mechanisms, independent of apoferritin synthesis, which increase the rate of iron absorption. Iron transferred to the capillaries is trapped exclusively by transferrin.

22.4 IRON DISTRIBUTION AND KINETICS

A normal 70-kg male has 3–4 g of iron, of which only 0.1% (3.5 mg) is in the plasma. Approximately 2.5 g are in hemoglobin. Table 22.1 lists the distribution of iron in the human.

Normally about 33% of the sites on transferrin contain iron. Iron

Table 22.1 Approximate Iron Distribution: 70–kg Man

	g	*%*
Hemoglobin	2.5	68
Myoglobin	0.15	4
Transferrin	0.003	0.1
Ferritin, tissue	1.0	27
Ferritin, serum	0.0001	0.004
Enzymes	0.02	0.6
Total	3.7	100

CLIN. CORR. 22.1 IRON-DEFICIENCY ANEMIA

Microscopic examination of a blood smear in iron-deficiency anemia usually reveals the characteristic findings of microcytic (small in size) and hypochromic (underpigmented) red blood cells. A bone marrow aspiration will reveal no storage iron to be present and serum ferritin values are virtually zero. The serum transferrin value (expressed as the total iron-binding capacity) will be elevated (upper limits of normal: 410 μg/dl) with a serum iron saturation of less than 16%. Common causes for iron deficiency include excessive menstrual flow, multiple births, and gastrointestinal bleeding which may be occult. The common causes of gastrointestinal bleeding include medication (especially aspirin or cortisone like drugs), hiatal hernia, peptic ulcer disease, gastritis associated with chronic alcoholism, and gastrointestinal tumors. The management of such patients must include both a careful examination for the cause and source of bleeding and also supplementation with iron. The latter is usually provided in the form of oral ferrous sulfate tablets; occasionally intravenous iron therapy may be required. Where the iron deficiency is severe, transfusion with packed red blood cells may also be indicated.

picked up from the intestine by transferrin is delivered primarily to the marrow for incorporation into the hemoglobin of red blood cells. The mobilization of iron from the mucosa and from storage sites involves in part the reduction of iron to the ferrous state and its reoxidation to the ferric form. The reduction mechanisms have not been well described. On the other hand, conversion of the ferrous ion back to the ferric state is regulated by serum enzymes called ferroxidases as indicated below.

$$Fe^{2+} + \text{ferroxidase} \longrightarrow Fe^{3+} + \text{reduced ferroxidase}$$

Ferroxidase I is also known as ceruloplasmin. A deficiency of this protein is associated with Wilson's disease, in which there is progressive hepatic failure and degeneration of the basal ganglia associated with a characteristic copper deposition in the cornea (Kayser–Fleischer rings). The ferroxidase activity of ceruloplasmin is not clinically important, since there is no evidence for significant impairment of mobilization of iron in Wilson's disease. Another serum protein, ferroxidase II, appears to be the major serum component that oxidizes ferrous ions. If an inappropriately large amount of iron is administered by injection to a subject who is not iron-deficient, this iron will be transported to the liver for storage in the form of ferritin.

In any disease process where iron loss exceeds iron repletion, a sequence of physiological responses occurs. The initial events are without symptoms to the subject and involve depletion of iron stores without compromise of any physiological function. This depletion will be manifested by a reduction or absence of iron stores in the liver and in the bone marrow and also by a decrease in the content of the very small amount of ferritin that is normally present in plasma. Serum ferritin levels reflect slow release from storage sites during the normal cellular turnover that occurs in the liver; measurements are made by radioimmune assays. Serum ferritin is mostly apoferritin in form, containing very little iron. During this early phase, the level and percent saturation of serum transferrin is not distinctly abnormal. As the iron deficiency progresses, the level of hemoglobin begins to fall and morphological changes appear in the red blood cells. Concurrently, the serum iron falls with a rise in the level of total serum transferrin, the latter reflecting a physiological adaptation in an attempt to absorb more iron from the gastrointestinal tract. At this state of iron depletion a very sensitive index is the percent saturation of serum transferrin with iron (normal range 21—50%). At this point the patient usually comes to medical attention and the diagnosis of iron deficiency is made. In countries where iron deficiency is severe without available corrective medical measures, a

third and severe stage of iron deficiency can occur, where there begins to be a depletion of iron-containing enzymes leading to very pronounced metabolic effects (see Clin. Corr. 22.1).

Iron overload can occur in patients so that the iron content of the body can be elevated to values of about 100 g. This may happen for a variety of reasons. Some patients have a recessive heritable disorder associated with a marked inappropriate increase in iron absorption. In such cases the serum transferrin can be almost completely saturated with iron. This state, which is known as idiopathic hemochromatosis, is more commonly seen in men because women with the abnormal gene are protected somewhat by menstrual and childbearing events. The accumulation of iron in the liver, pancreas, and the heart can lead, respectively, to cirrhosis and liver tumors, diabetes mellitus, and cardiac failure. The treatment for these patients is periodic withdrawals of large amounts of blood where the iron is in the contained hemoglobin. Another group of patients have severe anemias, among the most common of which are the thalassemias, a group of congenital hemolytic anemias. In these cases the subjects require a large amount of transfusions throughout their lives leading to the accumulation of large amounts of iron derived from the transfused blood. Clearly bleeding would be an inappropriate measure in these cases; rather, the patients are treated by the administration of iron chelators, such as desferrioxamine, which leads to the excretion of large amounts of complexed iron in the urine. Rarely, a third group of patients will acquire excess iron because they ingest large amounts of both iron and ethanol, the latter promoting iron absorption. In these cases excess stored iron can be removed by bleeding. (see Clin. Corr. 22.2).

CLIN. CORR. 22.2 HEMOCHROMATOSIS AND IRON-FORTIFIED DIET

For many years there was a controversy as to whether food should be fortified with iron because of the prevalence of iron-deficiency anemia, especially among premenopausal women. Proponents suggested that if at least 50 mg of iron was incorporated per pound of enriched flour, dietary iron deficiency would be reduced markedly. On the other hand, objecters stated that the possibility of toxicity from excess iron absorption through iron fortification was too great, and thus such a measure should not be sponsored. Recent studies have indicated that the gene prevalence for hemochromatosis is extraordinarily high, about 10% in the general population. Since the disease is expressed primarily in the homozygous state, ~1% of all individuals are at risk for hemochromatosis. A study in Sweden measured the serum iron and iron-binding capacity in a group of 347 subjects. No women showed evidence of iron overload. However, 5% of the men had persistent elevation of serum iron values with 2% of the men having increased iron stores consonant with the distribution found in early stages of hemochromatosis. Of relevance is the fact that Sweden has had mandated iron fortification for the past 30 years, and approximately 42% of the average daily intake of iron is derived from fortified sources. This study points out the danger of iron-fortified diets. In other settings, however, such as in countries where iron-deficiency is widespread, fortification may still be the most appropriate measure.

22.5 HEME BIOSYNTHESIS

Heme is produced in virtually all mammalian tissues. Its synthesis is most pronounced in the bone marrow and liver because of the requirements for incorporation into hemoglobin and the cytochromes, respectively. As depicted (see Figure 22.4), heme is largely a planar molecule. It consists of one ferrous ion and a tetrapyrrole ring, protoporphyrin IX. The diameter of the iron atom is a little too large to be accommodated within the plane of the porphyrin ring, and thus the metal puckers out to one side as it coordinates with the apical nitrogens of the four pyrrole groups. Heme is one of the most stable of compounds, reflecting its strong resonance features.

Figure 22.5 depicts the pathway for heme biosynthesis. The fol-

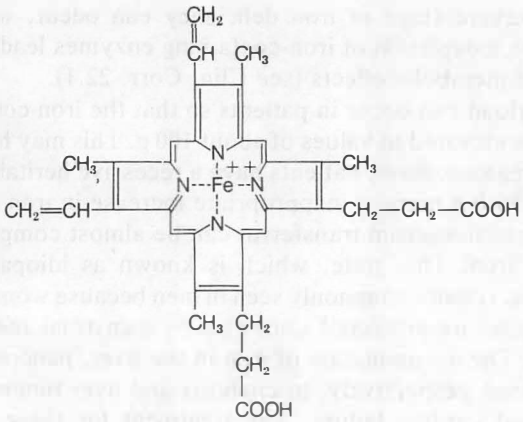


Figure 22.4
Structure of heme.

lowing are the important aspects to be noted. First, the initial and last three enzymatic steps are catalyzed by enzymes that are in the mitochondrion, whereas the intermediate steps take place in the cytoplasm. This is important in considering the regulation by heme of the first biosynthetic step; this aspect is discussed below. Second, the organic portion of heme is derived totally from eight residues each of glycine and succinyl CoA. Third, the reactions occurring on the side groups attached to the tetrapyrrole ring involve the colorless intermediates known as porphyrinogens. The latter compounds, though exhibiting resonance features within each pyrrole ring, do not demonstrate resonance between the pyrrole groups. As a consequence, the porphyrinogens are unstable and can be readily oxidized, especially in the presence of light, by nonenzymatic means to their stable porphyrin products. In the latter cases resonance between pyrrole groups is established by oxidation of the four methylene bridges. Figure 22.6 depicts the enzymatic conversion of protoporphyrinogen to protoporphyrin by this oxidation mechanism. This is the only known porphyrinogen oxidation that is enzyme-regulated in man; all other porphyrinogen-to-porphyrin conversions are nonenzymatically and light catalyzed rather than by specific enzymes. Fourth, once the tetrapyrrole ring is formed, the order of the R groups as one goes clockwise around the tetrapyrrole ring defines which of the four possible types of uro- or coprophyrinogens are being synthesized. These latter compounds have two different substituents, one each for every pyrrole group. Going clockwise around the ring, the substituents can be arranged as ABABABAB

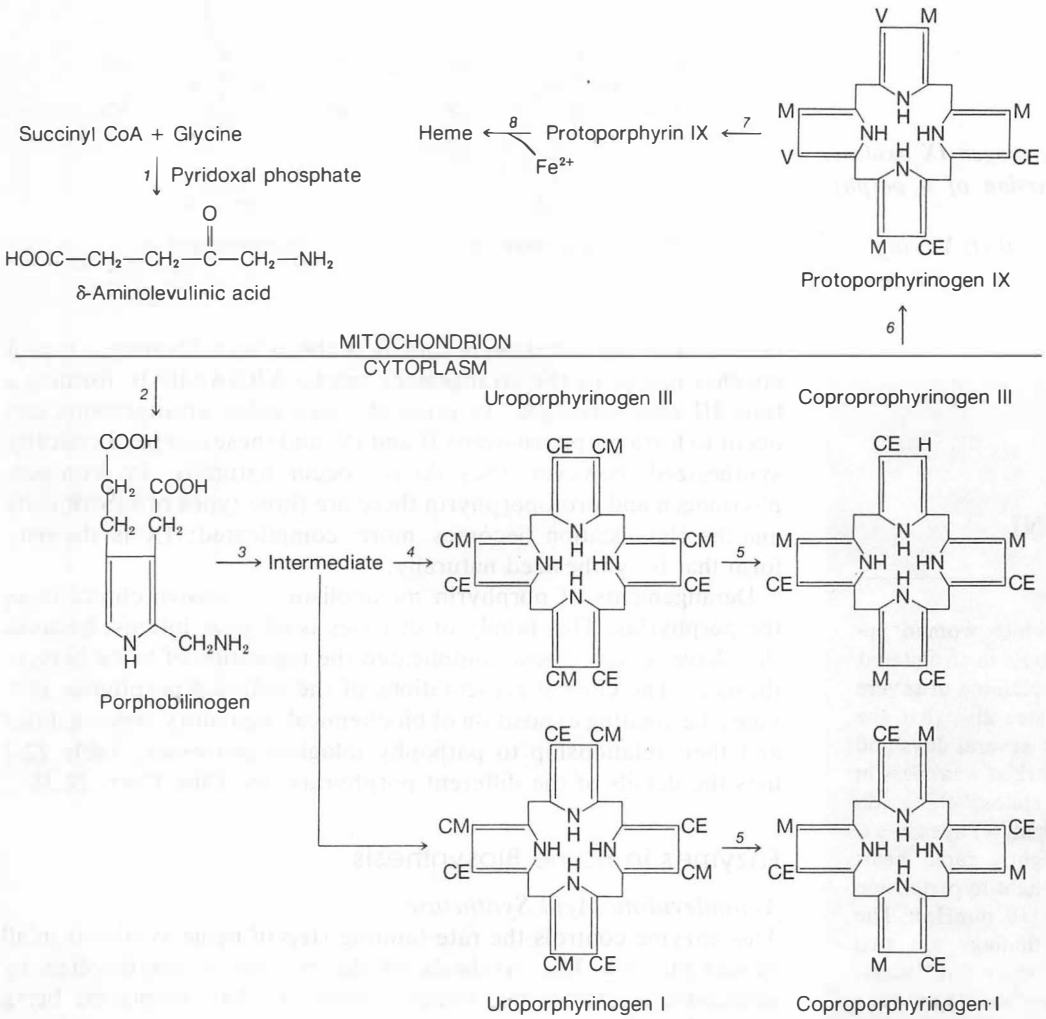


Figure 22.5
Pathway for heme biosynthesis.
The numbers indicate the proteins involved in each of the biochemical steps according to the following code: 1, Ala synthetase; 2, Ala dehydratase; 3, uroporphyrinogen I synthetase; 4, uroporphyrinogen III cosynthetase; 5, coproporphyrinogen decarboxylase; 6, coproporphyrinogen III oxidase; 7, protoporphyrinogen IX oxidase; 8, ferrochelatase. The pyrrole ligands are indicated by the following abbreviations: CE, β-carboxyethyl; CM, carboxymethyl; M, methyl; V, vinyl.

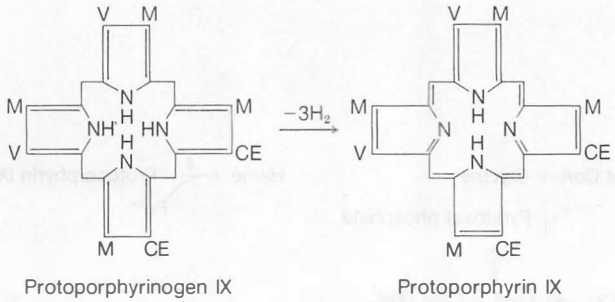


Figure 22.6
The action of protoporphyrinogen IX oxidase, an example of the conversion of a porphyrinogen to a porphyrin.
CE, β-carboxyethyl; M, methyl; V, vinyl.

CLIN. CORR. 22.3 ACUTE INTERMITTENT PORPHYRIA

A 40-year-old single, white woman appears in the emergency room in an agitated state, weeping, and complaining of severe abdominal pain. She states also that she has been constipated for several days and has noted a feeling of marked weakness in her arms and legs and states "things do not appear to be quite right." Physical examination reveals a slightly rapid heart rate (110/min) and also slight hypertension (blood pressure of 160/110 mmHg). The only other significant findings are two well-healed abdominal operative scars. When queried, she relates that there have been earlier episodes of severe abdominal pain, in fact more severe than what she is presently experiencing. Exploratory abdominal operations undertaken on two of those past occasions revealed no abnormalities.

The usual laboratory tests are obtained and appear to be largely within normal limits. None of her neurological complaints appear to be either well-documented or have any localized anatomical focus. The

(where A is one substituent and B is the other), forming a type I porphyrinogen, or the arrangement can be ABBAABAB, forming a type III porphyrinogen. In principle, two other arrangements can occur to form porphyrinogens II and IV, and these can be chemically synthesized; however, they do not occur naturally. In protoporphyrinogen and protoporphyrin there are three types of substituents and the classification becomes more complicated; IX is the only form that is synthesized naturally.

Derangements of porphyrin metabolism are known clinically as the porphyrias. This family of diseases is of great interest because they have revealed how complicated the regulation of heme biosynthesis is. The clinical presentations of the different porphyrias provide a fascinating exposition of biochemical regulatory abnormalities and their relationship to pathophysiological processes. Table 22.2 lists the details of the different porphyrias (see Clin. Corr. 22.3).

Enzymes in Heme Biosynthesis

Aminolevulinic Acid Synthetase

This enzyme controls the rate-limiting step of heme synthesis in all tissues studied. The synthesis of the enzyme is not directed by mitochondrial DNA but occurs rather in the cytoplasm, being directed by mRNA derived from the nucleus. The enzyme is incorporated into the outer membrane of the mitochondrion and functions at that site only. Succinyl CoA is one of the substrates and is found only in the mitochondrion. Recently this protein has been purified to apparent homogeneity from rat liver mitochondrion. The enzyme consists of a dimer of subunits of 60,000 daltons each; 50% inhibition of activity occurs in the presence of 5 μM hemin, and virtually complete inhibition is noted at a 20-μM concentration. The enzyme has a short biological half-life (about 60 min). Both the synthesis and the activity of the enzyme are subject to regulation by a variety of substances. The enzymatic reaction involves the condensation of a

Table 22.2 Derangements in Porphyrin Metabolism

Disease State	*Genetics*	*Tissue*	*Enzyme*	*Activity*	*Organ Pathology*
Acute intermittent porphyria	Dominant	Liver	1. ALA synthetase 2. Uroporphyrinogen I synthetase 3. Δ^4-5α-Reductase	Increase Decrease Decrease	Nervous system
Hereditary coproporphyria	Dominant	Liver	1. ALA synthetase 2. Coproporphyrinogen oxidase	Increase Decrease	Nervous system; skin
Variegate porphyria	Dominant	Liver	1. ALA synthetase 2. Protoporphyrinogen oxidase	Increase Decrease	Nervous system; skin
Porphyria cutanea tarda	Dominant	Liver	1. Uroporphyrinogen decarboxylase	Decrease	Skin, induced by liver disease
Hereditary protoporphyria	Dominant	Marrow	1. Ferrochelatase	Decrease	Gallstones, liver disease, skin
Erythropoietic porphyria	Recessive	Marrow	1. Uroporphyrinogen III cosynthetase	Decrease	Skin and appendages; reticuloendothelial system
Lead poisoning	None	All tissues	1. ALA dehydrase 2. Ferrochelatase	Decrease Decrease	Nervous system; blood; others

glycine residue with a residue of succinyl CoA (Figure 22.7). The reaction has an absolute requirement for pyridoxal phosphate; the latter interacts with the nitrogen of glycine to form a Schiff's base. This generates a carbanion intermediate on the α-carbon of the glycine, allowing a nucleophilic attack and condensation with the succinyl group from succinyl CoA. The bound intermediate, α-amino-β-ketoadipic acid, decarboxylates to form the released product, aminolevulinic acid (ALA). Pyridoxal deficiencies and drugs competing with pyridoxal lead to a decrease in enzyme activity.

ALA Dehydrase

This enzyme is a soluble cytosol component with a molecular weight of 280,000 and consists of eight subunits, of which only four interact with the substrate. This protein also interacts with the substrate to form a Schiff's base, but in this case the ε-amino group of a lysine residue binds to the ketonic carbon of one substrate molecule (Figure 22.8). Two molecules of ALA condense asymmetrically to form porphobilinogen. ALA dehydrase is a sulfhydryl enzyme and is very sensitive to inhibition by heavy metals. A characteristic finding of

decision is made that the present symptoms are largely psychiatric in origin and are of a functional rather than an organic basis. Because of her agitated state, the decision is made to sedate the patient with 60 mg of phenobarbital; a consultant psychiatrist agrees by telephone to see the patient in about 4 hours. During the ensuing interval, the emergency room staff notices marked deterioration in the patient's status: generalized weakness rapidly appears, progressing to a compromise of respiratory function. This ominous development leads to immediate incorporation of a ventilatory assistance regimen, with transfer of the patient to the intensive care unit for close physiological monitoring. Despite these measures the patient's condition deteriorates further and she dies 48 hours later. Shortly before death a urine

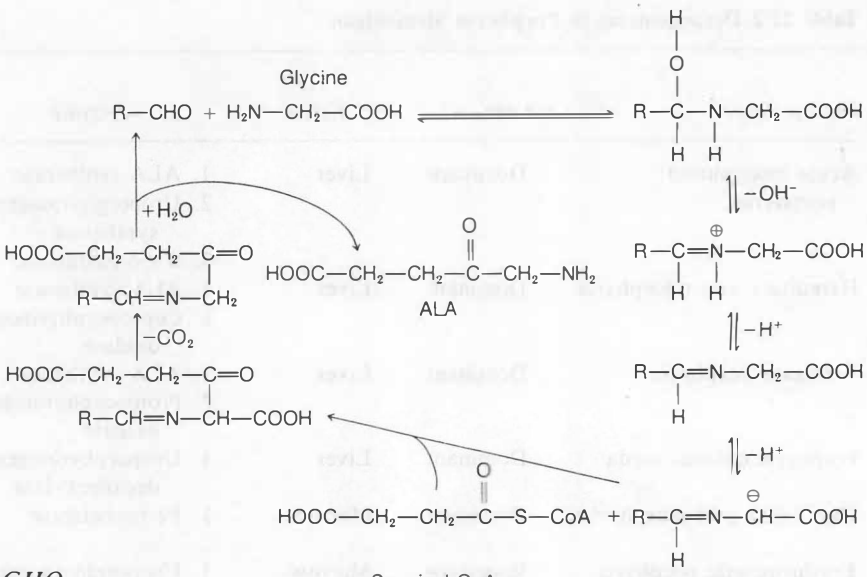


Figure 22.7
The synthesis of δ-aminolevulinic acid.
Pyridoxal phosphate is indicated by R—CHO.

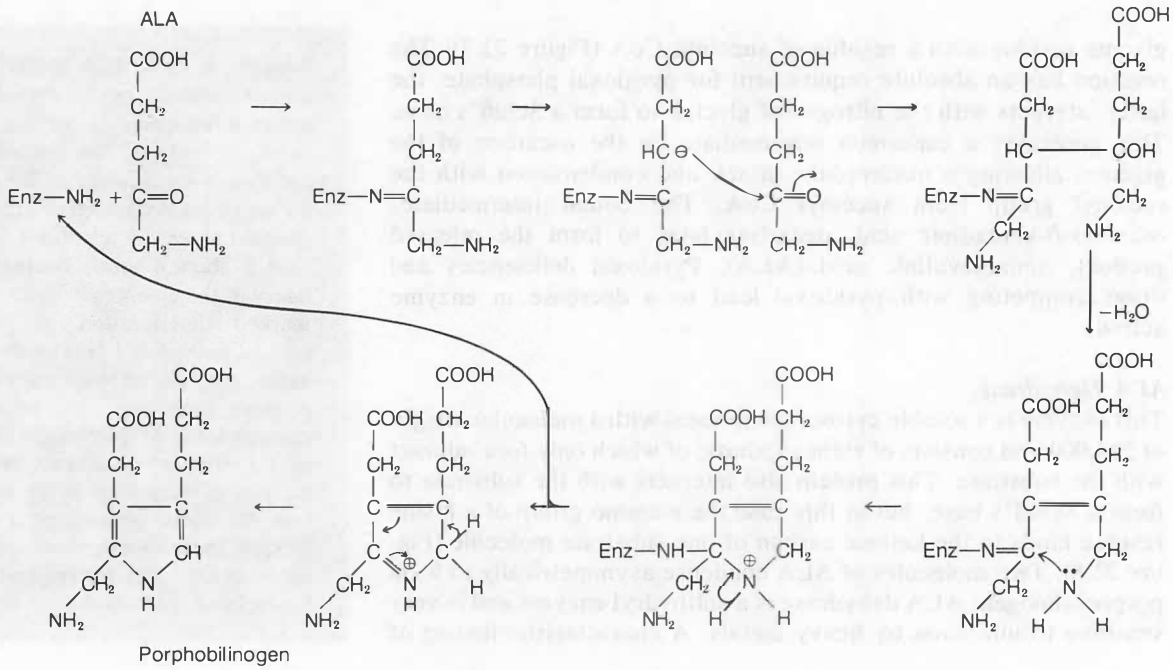


Figure 22.8
The synthesis of porphobilinogen.

lead poisoning is the elevation of ALA in the absence of an elevation of porphobilinogen.

Uroporphyrinogen I Synthetase

The synthesis of uroporphyrinogen I involves head-to-tail condensations of four porphobilinogen residues, with the deamination of the primary amino groups. This enzyme has recently been purified from red blood cells and has been found to consist of a multiplicity of isoenzymes. The mechanism and the nature of the intermediates in this multistep reaction are not known. Normally this enzyme is closely associated with a second protein, *uroporphyrinogen III cosynthetase.* The latter protein does not have any apparent enzymatic function but acts rather as a specifier protein, directing uroporphyrinogen I synthetase to form the III isomer, rather than the I isomer of uroporphyrinogen. The nature of the complicated biochemical steps modulated by the cosynthetase is as yet undetermined. A rare recessive disease, erythropoietic porphyria, associated with marked light sensitization, is due to an abnormality of red cell uroporphyrinogen III cosynthetase. Here large amounts of the type I isomers of uroporphyrinogen and coproporphyrinogen are synthesized.

Uroporphyrinogen Decarboxylase

This enzyme acts on the side chains of the uroporphyrinogens to form the coproporphyrinogens. The protein catalyzes the conversion of both I and III isomers of uroporphyrinogen to the respective coproporphyrinogen isomers. Uroporphyrinogen decarboxylase is inhibited by iron salts. Clinically the most common cause of porphyrin derangement is associated with patients who have a single gene abnormality for this enzyme, leading to 50% depression of the enzyme's activity. This disease, which shows primarily cutaneous manifestations with sensitivity to light, is known as porphyria cutanea tarda. The condition is not expressed unless patients either take drugs that cause an increase in porphyrin synthesis or drink large amounts of alcohol, leading to the accumulation of iron, which then acts to inhibit further the activity of uroporphyrinogen decarboxylase.

Coproporphyrinogen Oxidase

This mitochondrial enzyme is specific for the type III isomer of coproporphyrinogen, not acting on the type I isomer. Coproporphyrinogen III enters the mitochondrion and is converted to protoporphyrinogen IX. The mechanism of action is not understood. A dominant disease associated with a deficiency of this enzyme leads to a form of hereditary hepatic porphyria, known as hereditary coproporphyria.

sample of the patient is found to have a markedly elevated level of porphobilinogen.

This patient had acute intermittent porphyria, a disease of incompletely understood derangement of heme biosynthesis. This entity must be considered in any postpubertal patient who develops unexplained neurological, abdominal, or psychiatric symptoms. There is a dominant pattern of inheritance associated with an overproduction of the porphyrin precursors, ALA and porphobilinogen. Several associated well-defined enzyme abnormalities are noted in the cases that have been studied carefully. These include (1) a marked increase in ALA synthetase, (2) a reduction by half of the activity of uroporphyrinogen I synthetase, and (3) a reduction of one-half of the activity of steroid Δ^4-5α-reductase. The change in content of the second enzyme is consonant with a dominant expression. The change in content of the third enzyme is acquired and not apparently a heritable expression of the disease. It is believed that a decrease in uroporphyrinogen I synthetase leads to a minor decrement in the content of heme in the liver. The low concentration of heme leads to both a failure to repress the synthesis and to inhibit the activity of ALA synthetase. Since this disease is almost never manifested before puberty, it is thought that only with the induction of Δ^4-5β-reductase at adolescence does the disease become florid. Since these patients do not have a sufficient amount of Δ^4-5α-reductase, it is assumed that the observed increase in the 5β steroids is due to a shunting of Δ^4 steroids into the 5β-reductase pathway.

Pathophysiologically, the disease poses a great riddle: the derangement of porphyrin metabolism is confined to the liver, which anatomically appears normal, whereas the pathological findings are restricted to the nervous system. In the present case, involvement of (1) the brain led to the agitated and confused state, (2) the

autonomic system led to the hypertension, increased heart rate, and constipation with secondary abdominal pain, and (3) the peripheral nervous system and spinal cord, led to the weakness and sensory disturbances. The conclusion generated by these observations is that there must be some hepatic product that circulates to reach and affect neural tissue. However, experimentally no known intermediate of heme biosynthesis can cause the pathology noted in acute intermittent porphyria. In the present case there should have been a greater suspicion of the possibility of porphyria early in the patient's presentation. The analysis for porphobilinogen in the urine is a relatively simple test. The treatment would have been glucose infusion, the exclusion of any drugs that could cause elevation of ALA synthetase and, if her disease failed to respond satisfactorily despite these measures, the administration of intravenous hematin to inhibit the synthesis and activity of ALA synthetase.

Acute intermittent porphyria is of historic political interest. The disease has been diagnosed in two descendents of George III, suggesting that the latter's deranged personality preceding and during the American Revolution could be ascribed to an affliction with this illness.

Protoporphyrinogen Oxidase

This mitochondrial enzyme generates a product, protoporphyrin IX, which, in contrast to the other heme precursors, is very water-insoluble. Excess amounts of protoporphyrin IX that are not converted to heme are excreted by the biliary system into the intestinal tract. A dominant disease, variegate porphyria, is probably due to a deficiency of protoporphyrinogen oxidase.

Ferrochelatase

This enzyme inserts ferrous iron into protoporphyrin IX in the final step of the synthesis of heme. Reducing substances are required for its activity. The protein is sensitive to the effects of heavy metals (especially lead) and, of course, to iron deprivation. In these latter instances, zinc instead of iron is incorporated to form a zinc–protoporphyrin IX complex. In contrast to heme, the zinc–protoporphyrin IX complex is brilliantly fluorescent and easily detectable in small amounts.

Regulation of Heme Biosynthesis

ALA synthetase controls the rate-limiting step of heme synthesis in all tissues. Succinyl CoA and glycine are substrates for a variety of reactions. The modulation of the activity of ALA synthetase determines the quantity of the substrates that will be shunted into heme biosynthesis. Heme (and also hematin) acts both as a repressor of the synthesis of ALA synthetase and also as an inhibitor of its activity. Since heme resembles neither the substrates nor the product of the enzyme's action, it is probable that the inhibition occurs at an allosteric site. Almost 100 different drugs and metabolites can cause induction of ALA synthetase; for instance, a 40-fold increase is noted in the rat after treatment with 3,5-dicarbethoxy-1,4-dihydrocollidine. The effect of pharmacological agents has led to the important clinical feature wherein some patients with certain kinds of porphyria have had exacerbations of their condition following the inappropriate administration of certain drugs (e.g., barbiturates). ALA dehydrase is also inhibited by heme; but this is of little physiological consequence, since the maximal velocity of the total amount of ALA dehydrase present is about 80-fold greater than that of ALA synthetase, and thus heme-inhibitory effects are reflected first in the activity of ALA synthetase.

Glucose or one of its proximal metabolites serves to inhibit heme biosynthesis in a mechanism which is not yet defined. This is of clinical relevance, since some patients manifest their porphyric state for the first time when placed on a very low caloric (and therefore glucose) intake. Other regulators of porphyrin metabolism include

certain steroids. Steroid hormones (e.g., oral contraceptive pills) with a double bond in ring A between carbon atoms 4 and 5 can be reduced by two different reductases. The product of 5α-reduction has little effect on heme biosynthesis; however, the product of 5β-reduction serves as a stimulus for the synthesis of ALA synthetase. The observation that 5β-reductase appears at puberty suggests this to be the reason why in a few of the porphyrias manifestations are not present at an earlier stage.

22.6 HEME CATABOLISM

The catabolism of heme-containing proteins presents two requirements to the mammalian host: (1) the development of a means of processing the hydrophobic products of porphyrin ring cleavage, and (2) the retention and mobilization of the contained iron so that it may be reutilized.

The major site of heme catabolism takes place within the red blood cells, the life span of which is approximately 120 days. Senescent cells are recognized by their membrane changes and removed and engulfed by the reticuloendothelial system at extravascular sites. The globin chains denature, releasing heme into the cytoplasm. The globin is degraded to its constituent amino acids, which are reutilized for general metabolic needs.

Figure 22.9 depicts the sequence of events of heme catabolism. Heme is degraded primarily by a microsomal enzyme system in reticuloendothelial cells which requires molecular oxygen and NADPH. Cytochrome c serves as the major vehicle for regenerating the NADPH utilized in the reaction. Heme oxygenase is substrate-inducible. The enzyme specifically catalyzes the cleavage of the α-methene bridge, which joins the two pyrrole residues containing the vinyl substituents. The α-methene carbon is converted quantitatively to carbon monoxide. The only endogenous source of carbon monoxide in man is the α-methene carbon. A fraction of the carbon monoxide is released via the respiratory tract. Thus the measurement of carbon monoxide in an exhaled breath provides an index to the quantity of heme that is degraded in an individual. The oxygen present in the carbon monoxide and the newly derivatized lactam rings are derived entirely from molecular oxygen. The stoichiometry of the reaction requires 3 mol oxygen for each ring cleavage. Heme oxygenase will only use heme as a substrate with the iron possibly participating in the cleavage mechanism. Thus, free protoporphyrin

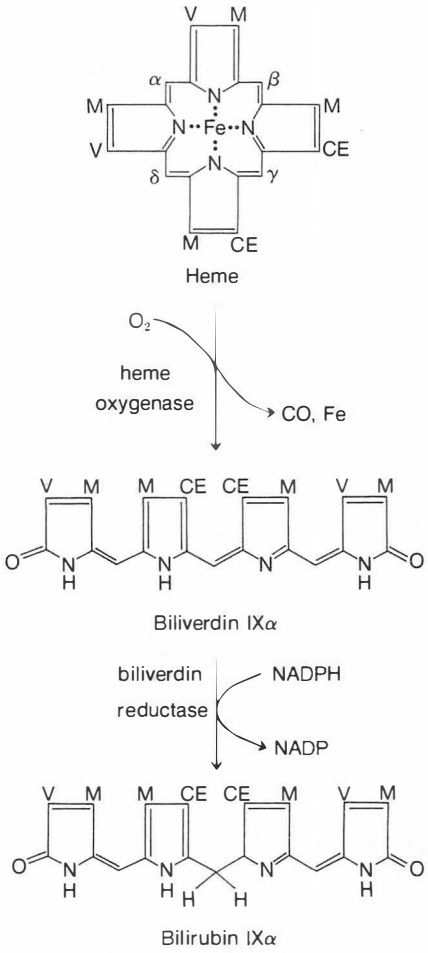


Figure 22.9
The formation of bilirubin from heme.
The Greek letters indicate the nomenclature for the methene carbons in heme. CE, β-carboxyethyl; M, methyl; V, vinyl.

IX is not a substrate. The linear tetrapyrrole biliverdin IX is the product formed by the action of heme oxygenase. Biliverdin IX is reduced by biliverdin reductase to bilirubin IX.

Bilirubin Metabolism

Bilirubin is derived not only from senescent red cells but also from the turnover of other heme-containing proteins, such as the cytochromes. Studies with labeled glycine as a precursor have revealed that an early-labeled bilirubin, with a peak amount present within 1–3 h, appears very shortly after a pulsed administration of the labeled precursor. A larger amount of bilirubin appears much later at around 120 days, reflecting the turnover of heme in red blood cells. Early-labeled bilirubin can be divided into two parts: an early-early part, which reflects the turnover of heme proteins in the liver, and a late-early part, which consists of both the turnover of heme-containing hepatic proteins and also the turnover of bone marrow heme, which is either poorly incorporated or easily released from red blood cells. The latter is a measurement of ineffective erythropoiesis and can be very pronounced in such disease states as pernicious anemia (see Chapter 26) and the thalassemias.

Bilirubin is poorly soluble in aqueous solutions at physiological pH values. When transported in plasma, it is bound to serum albumin with an association constant greater than 10^6 M^{-1}. Albumin contains one such high affinity site and another with a lesser affinity. At the normal albumin concentration of 4 g/dl, ~70 mg of bilirubin/dl plasma can be bound on the two sites. However, bilirubin toxicity (kernicterus), which is manifested by the transfer of bilirubin to membrane lipids, commonly occurs at concentrations >25 mg/dl. This suggests that the weak affinity of the second site does not allow it to serve effectively in the transport of bilirubin. Bilirubin on serum albumin is rapidly cleared by the liver, where there is a free bidirectional flux of the tetrapyrrole across the sinusoidal–hepatocyte interface. Once in the hepatocyte, bilirubin is bound to several cytosol proteins, of which only one has been well characterized. The latter component, ligandin, is a small basic protein, making up to 6% of the total cytoplasmic protein of rat liver. Recently ligandin has been purified to homogeneity from rat liver and characterized as having two subunits of molecular weight 22,000 and 27,000. Each subunit contains glutathione *S*-epoxide transferase activity, a function important in detoxification mechanisms of aryl groups. The stoichiometry of binding is one bilirubin residue per complete ligandin molecule. The functional role of ligandin and other hepatic bilirubin-binding proteins remains to be defined.

Once in the hepatocyte the β-carboxyethyl side chains of bilirubin are conjugated to form a diglucuronide (see Figure 22.10). The reaction mechanism includes the utilization of uridine diphosphoglucose, which is oxidized by a dehydrogenase to uridine diphosphoglucuronate. The latter serves as a glucuronate donor to bilirubin through the action of a specific transferase, which forms sequentially the mono- and diglucuronide adducts of bilirubin. The mammalian liver contains several uridine diphosphoglucuronyltransferases, each of which is substrate-specific for the acceptor molecule. In normal bile the diglucuronide is the major form of excreted bilirubin, with only small amounts present of the monoglucuronide or other glycosidic adducts. Bilirubin diglucuronide is much more water-soluble than free bilirubin, and thus the transferase facilitates the excretion of the bilirubin into bile. Bilirubin diglucuronide is poorly absorbed by the

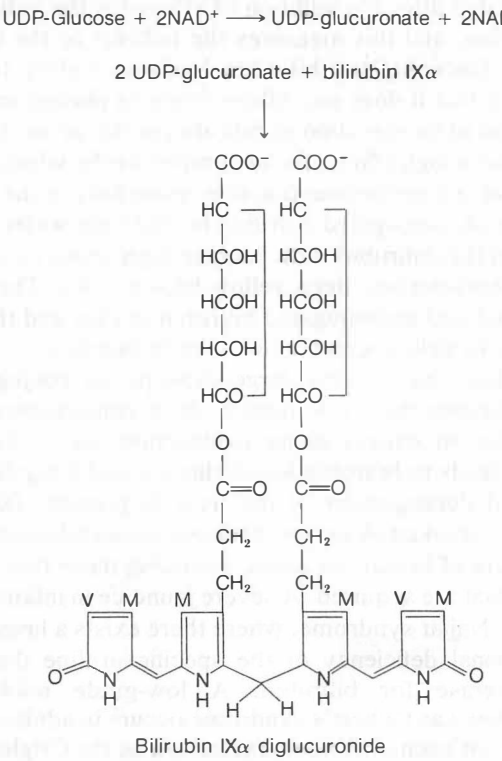


Figure 22.10
The biosynthesis of bilirubin diglucuronide.
M, methyl; V, vinyl.

intestinal mucosa. The glucuronide residues are released in the terminal ileum and large intestine by bacterial hydrolases, and the released free bilirubin is reduced to the colorless linear tetrapyrroles known as urobilinogens. Urobilinogens can be oxidized to colored products known as urobilins, which are excreted in the feces. A small fraction of urobilinogen can be reabsorbed by the terminal ileum and large intestine to be removed by hepatic cells and resecreted in bile. When urobilinogen is reabsorbed in large amounts in certain disease states, the kidney serves as a major excretory site.

In the normal state plasma bilirubin concentrations are 0.3–1 mg/dl, and this is almost all in the unconjugated state. In the clinical setting conjugated bilirubin is expressed as direct bilirubin because it can be coupled readily with diazonium salts to yield azo dyes; this is the direct van den Bergh reaction. Unconjugated bilirubin is bound to albumin and will not react until it is released by the addition of an organic solvent such as ethanol. The reaction with diazonium salts yielding the azo dye after the addition of ethanol is the indirect van den Bergh reaction, and this measures the indirect or the unconjugated bilirubin. Unconjugated bilirubin binds so tightly to serum albumin and lipid that it does not diffuse freely in plasma and therefore does not lead to an elevation of bilirubin in the urine. Unconjugated bilirubin has a high affinity for membrane lipids, which leads to the impairment of cell membrane function, especially in the nervous system. In contrast, conjugated bilirubin is relatively water-soluble, and elevations of this bilirubin form leads to high urinary concentrations with the characteristic deep yellow-brown color. The deposition of conjugated and unconjugated bilirubin in skin and the sclera gives the yellow to yellow-green color seen in jaundice.

The normal liver has a very large capacity to conjugate and mobilize the bilirubin that is delivered. As a consequence hyperbilirubinemia due to excess heme destruction, as in hemolytic diseases, rarely leads to bilirubin levels that exceed 4 mg/dl, except where functional derangement of the liver is present. (See Clin. Corr. 22.4.) Thus, marked elevation of unconjugated bilirubin reflects primarily a variety of hepatic diseases, including those that are heritable and those that are acquired. A severe jaundice in infants occurs with the Crigler–Najjar syndrome, where there exists a homozygous complete functional deficiency of the specific uridine diphosphoglucuronyltransferase for bilirubin. A low-grade mild hyperbilirubinemia known as Gilbert's syndrome occurs in adults. Though this disease has not been as well characterized as the Crigler–Najjar syndrome, one of the findings is a moderate reduction in bilirubin-uridine diphosphoglucuronyltransferase activity.

Elevations of conjugated bilirubin levels in plasma are attributable to liver and/or biliary tract disease. In simple uncomplicated biliary

CLIN. CORR. 22.4 NEONATAL ISOIMMUNE HEMOLYSIS

Rh-negative women pregnant with Rh-positive fetuses will develop antibodies to Rh factors. These antibodies will cross the placenta to hemolyze fetal red blood cells. Usually this is not of clinical relevance until about the third Rh-positive pregnancy, where the mother has had antigenic

tract obstruction, the major component of the elevated serum bilirubin is the diglucuronide form, which is released by the liver into the vascular compartment. Biliary tract disease may be extrahepatic or may be intrahepatic, the latter involving the canaliculi and biliary ductules. The Dubin–Johnson syndrome is an autosomal recessive disease involving a defect in the biliary secretory mechanism of the liver. The excretion through the biliary tract of a variety of (but not all) organic anions is affected. The retention of undefined pigments in the liver in this disorder leads to a characteristic gray-black color of this organ. This pigment is not apparently derived from heme. A second heritable disorder associated with elevated levels of plasma conjugated bilirubin is Rotor syndrome. In this poorly defined disease no hepatic pigmentation occurs, and the associated finding of increased secretion of urinary coproporphyrins I and III are found. The relationship of the derangement of coproporphyrin metabolism to bilirubin metabolism is not well understood.

Intravascular Hemolysis

In certain diseases destruction of red blood cells occurs in the intravascular compartment rather than in the extravascular reticuloendothelial cells. In the former case the appearance of free hemoglobin and heme in the plasma could lead potentially to the excretion of these substances through the kidney with a substantial loss of iron. To prevent this occurrence specific plasma proteins are involved in scavenging mechanisms. Transferrin binds free iron and thus permits the reutilization of the metal. Free hemoglobin in the plasma leads to the following sequence of events. After oxygenation in the pulmonary capillaries, plasma oxyhemoglobin dissociates into $\alpha\beta$ dimers, which are bound to a family of circulating plasma proteins, the haptoglobins, having a high affinity for the oxyhemoglobin dimer. Since deoxyhemoglobin does not dissociate into dimers in physiological settings, it is not bound by haptoglobin. The stoichoimetry of binding is two $\alpha\beta$-oxyhemoglobin dimers per haptoglobin molecule. Interesting studies have been made with rabbit antihuman-hemoglobin antibodies on the haptoglobin–hemoglobin interaction. Human haptoglobin interacts with a variety of hemoglobins from different species. The binding of human haptoglobin with human hemoglobin is not affected by the binding of rabbit antihuman-hemoglobin antibody. These studies suggest that haptoglobin binds to sites on hemoglobin, which are highly conserved in evolution and therefore are not sufficiently antigenic to generate antibodies. The most likely site for the molecular interaction of hemoglobin and haptoglobin is the interface of the α and β chains of the tetramer which dissociates to yield the $\alpha\beta$ dimer. Sequence determinations with earlier babies. Antenatal studies will reveal rising maternal titers of IgG antibodies against Rh-positive red blood cells, indicating that the fetus is Rh-positive. At birth these infants usually appear unremarkable; however, the unconjugated bilirubin in the umbilical cord blood is elevated up to 4 mg/dl, due to the hemolysis initiated by maternal antibodies. During the next 2 days the serum bilirubin rises, reflecting continuing isoimmune hemolysis. Prior to birth, placental transfer of fetal bilirubin occurs with excretion through the maternal liver. Because hepatic enzymes of bilirubin metabolism can be poorly expressed in the newborn, infants may not be able after delivery to process large amounts of bilirubin that can be generated. Treatment involves exchange transfusion with whole blood, which is serologically compatible with both the infant's blood and maternal serum. The latter requirement is necessary to prevent hemolysis of the transfused cells. Additional treatment includes external phototherapy, which facilitates the metabolism of bilirubin. The consequences of elevated plasma direct bilirubin include jaundice, hepatosplenomegaly, ascites, and edema. The signs of central nervous system damage with the appearance of lethargy, hypotonia, spasticity, and respiratory difficulty constitute the syndrome known as kernicterus.

minations have indicated that these contact regions are highly conserved in evolution.

The haptoglobins are α_2-globulins and are made in the liver. They consist of two pairs of polypeptide chains (α being the lighter and β the heavier). The β chains have a molecular weight of 42,600 and are invariant in structure; α chains are of three kinds. The shorter α^1 chains each consist of 84 residues, with a molecular weight of 9,000 varying only in the residues at positions 54 where Glu is found in α^{1S} and Lys in α^{1F}. The α^2 chain consists of an incomplete fusion-duplication of the genes for α^{1F} and α^{1S} leading to a polypeptide chain with 143 residues. The haptoglobin peptide chains are joined by disulfide bonds between the α and β chains and between the two α chains. Thus, in contrast to the immunoglobulins, the disulfide bond between similar chains occurs with the smaller polypeptides. Because the α^2 chain has one more half-cystine residue than the α^1 chain, haptoglobins with the formula $\alpha_2^2\beta_2$ can polymerize into larger aggregates through disulfide linkages. There may not be symmetry in any single haptoglobin molecule, thus haptoglobins are known which have the following molecular structures: $\alpha^{1F}\alpha^{1S}\beta_2$ or $\alpha^{1F}\alpha^2\beta_2$. These variations in the structure of haptoglobins have been useful in analyses of population genetics.

The interaction of haptoglobin with hemoglobin leads to a complex that is too large to be filtered through the renal glomerulus. Free hemoglobin appearing in renal tubules and in the urine will occur during intravascular hemolysis only when the binding capacity of circulating haptoglobin has been exceeded. Haptoglobin delivers hemoglobin to the reticuloendothelial cells. The heme in free hemoglobin is relatively resistant to the action of heme oxygenase, whereas the heme residues in an $\alpha\beta$ dimer of hemoglobin when attached to haptoglobin are very susceptible. This enhancement of oxygenase activity is especially pronounced with the $\alpha_2^2\beta_2$ haptoglobin type. It has been suggested that the high proportion of haptoglobin $\alpha_2^2\beta_2$ in certain populations (such as in southeast Asia) where there is a high incidence of hemolytic disease is a reflection of a selective genetic advantage. The measurement of serum haptoglobin is used clinically as an indication of the degree of intravascular hemolysis. Patients who have significant intravascular hemolysis will have low or absent levels of haptoglobin because of the removal of haptoglobin–hemoglobin complexes by the reticuloendothelial system. Haptoglobin levels can also be low in severe extravascular hemolysis, where the large load of hemoglobin in the reticuloendothelial system leads to the transfer of free hemoglobin into plasma.

Free heme and hematin appearing in plasma are bound by a β-globulin, hemopexin, which has a molecular weight of 57,000. One heme residue binds per hemopexin molecule. Hemopexin transfers

heme to the liver, where further metabolism by heme oxygenase occurs. In the normal state hemopexin contains very little bound heme, whereas in intravascular hemolysis, the hemopexin is almost completely saturated by heme and is cleared with a half-life of about 7 h. In the latter instance excess heme will bind to albumin, with newly synthesized hemopexin serving as a mediator for the transfer of the heme from albumin to the liver. Hemopexin also binds free protoporphyrin.

BIBLIOGRAPHY

Meyer, U. A., and Schmid, R. The porphyrias, in J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson (eds.), *The metabolic basis of inherited disease*. New York: McGraw-Hill, 1978, p. 1167.

Pollycove, M. Hemochromatosis, in J. B. Stanbury, J. B. Wyngaarden, and D. S. Frederickson (eds.). *The metabolic basis of inherited disease*. New York: McGraw-Hill, 1978, p. 1127.

Schmid, R., and McDonagh, A. F. Hyperbilirubinemia, in J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson (eds.). *The metabolic basis of inherited disease*. New York: McGraw-Hill, 1978, p. 1221.

23

Gas Transport and pH Regulation

JAMES BAGGOTT

23.1 INTRODUCTION TO GAS TRANSPORT

Large organisms, especially terrestrial ones, require a relatively tough, impermeable outer covering to help ward off dust, twigs, nonisotonic fluids like rain and seawater, and other elements of the environment that would be harmful to living cells. One of the consequences of being large and having an impermeable covering is that individual cells of the organism cannot exchange gases directly with the atmosphere. Instead there must exist a specialized exchange surface, such as a lung or a gill, and a system to circulate the gases (and other materials, such as nutrients and waste products) in a manner that will meet the needs of every living cell in the body.

The existence of a system for the transport of gases from the atmosphere to cells deep within the body is not merely necessary, it has definite advantages. Oxygen is a good oxidizing agent, and at its partial pressure in the atmosphere, about 160 mmHg or 21.3 kPa [1 mmHg = 1 torr = 0.133 kPa (kilopascal)], it would oxidize and inactivate many of the components of the cells, such as essential sulfhydryl groups of enzymes. By the time oxygen gets through the transport system of the body its partial pressure is reduced to a much less damaging 20 mmHg (2.67 kPa) or less. Similarly, carbon dioxide is relatively concentrated in the body and becomes diluted in transit to the atmosphere. In the tissues, where it is produced, its partial pressure is 46 mmHg (6.13 kPa) or more. In the lungs it is 40 mmHg (5.33 kPa), and in the atmosphere only 0.2 mmHg (0.03 kPa), less abundant than the rare gas argon. Its relatively high concentration in the body permits it to be used as one component of a physiologically important buffering system, a system that is particularly useful because, upon demand, the concentration of carbon dioxide in the extracellular fluid can be varied over a rather wide range. This is discussed in more detail later in this chapter.

Oxygen and carbon dioxide are carried between the lungs and the other tissues by the blood. In the blood some of each gas is present in simple physical solution, but mostly each is involved in some sort of interaction with hemoglobin, the major protein of the red blood cell. There is a reciprocal relation between hemoglobin's affinity for oxygen and carbon dioxide, so that the relatively high level of oxygen in the lungs aids the release of carbon dioxide, which is to be expired, and the high carbon dioxide level in other tissues aids the release of oxygen for use by those tissues. Thus, a description of the physiological transport of oxygen and carbon dioxide is the story of the interaction of these two compounds with hemoglobin.

23.2 NEED FOR A CARRIER OF OXYGEN IN THE BLOOD

An oxygen carrier is needed in the blood simply because oxygen is not soluble enough in blood plasma to meet the body's needs. At 38°C 1 liter of plasma will dissolve only 2.3 ml oxygen. Whole blood, with its hemoglobin, has a much greater oxygen capacity. One liter of blood normally contains about 150 g hemoglobin, and each gram of hemoglobin can combine with 1.34 ml oxygen. Thus the hemoglobin in 1 liter of blood can carry 200 ml of oxygen, 87 times as much as plasma alone would carry. Without an oxygen carrier, the blood would have to circulate 87 times as fast to provide the same capacity to deliver oxygen. As it is, the blood makes a complete circuit of the body in 60 s under resting conditions, and in the aorta it flows at the rate of about 18.6 m/s. An 87-fold faster flow would require a fabulous high-pressure pump, would produce tremendously turbulent flow and high shear forces in the plasma, would result in uncontrollable bleeding from wounds, and would not even allow the blood enough time in the lungs to take up oxygen. The availability of a carrier not only permits us to avoid these impracticalities, but also gives us a way of controlling oxygen transport, since the oxygen affinity of the carrier is responsive to changing physiological conditions.

The respiratory system includes the trachea, in the neck, which bifurcates in the thorax into right and left bronchi, as shown schematically in Figure 23.1. The bronchi continue to bifurcate into smaller and smaller passages, ending with tiny bronchioles, which open into microscopic gas-filled sacs called alveoli. It is in the alveoli that gas exchange takes place with the alveolar capillary blood.

As we inhale and exhale, the alveoli do not appreciably change in size. Rather, it is the airways that change in length and diameter as the air is pumped into and out of the lungs. Gas exchange between the airways and the alveoli then proceeds simply by diffusion. These anatomical and physiological facts have two important consequences. In the first place, since the alveoli are at the ends of long tubes that constitute a large dead space, and the gases in the alveoli are not completely replaced by fresh air with each breath, the gas composition of the alveolar air differs from that of the atmosphere, as shown in Table 23.1. Oxygen is lower in the alveoli because it is removed by the blood. Carbon dioxide is higher because it is added. Since we do not usually breathe air that is saturated with water vapor at 38°C, water vapor is generally added in the airways. The level of nitrogen is lower in the alveoli, not because it is taken up by

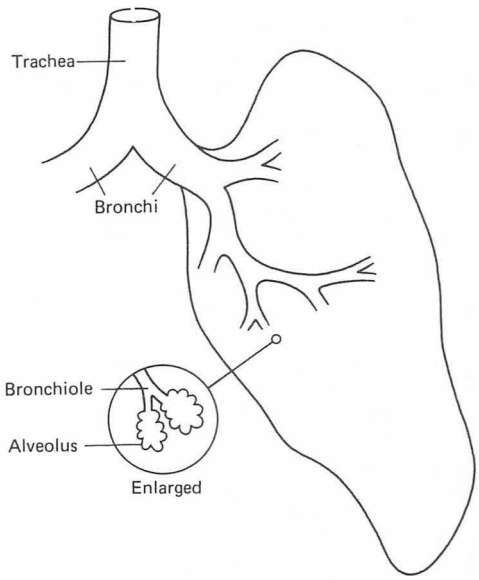


Figure 23.1
Diagram showing the respiratory tract.

Table 23.1 Partial Pressures of Important Gases Given in mmHg (kPa)

Gas	*In the Atmosphere*		*In the Alveoli of the Lungs*	
O_2	159	(21.2)	100	(13.3)
N_2	601	(80.1)	573	(76.4)
CO_2	0.2	(0.027)	40	(5.33)
H_2O	0	(0)	47	(6.27)
Total	760	(101)	760	(101)

the body, but simply because it is diluted by the carbon dioxide and water vapor.

A second consequence of the existence of alveoli of essentially constant size is that the blood that flows through the pulmonary capillaries during expiration, as well as the blood that flows through during inspiration, can exchange gases. This would not be possible if the alveoli collapsed during expiration and contained no gases, in which case the composition of the blood gases would fluctuate widely, depending on whether the blood passed through the lungs during an inspiratory or expiratory phase of respiration.

We have seen that an oxygen carrier is necessary. Clearly this carrier would have to be able to bind oxygen at an oxygen tension of about 100 mmHg (13.3 kPa), the partial pressure of oxygen in the alveoli. The carrier must also be able to release oxygen to the extrapulmonary tissues. The oxygen tension in the capillary bed of an active muscle is about 20 mmHg (2.67 kPa). In resting muscle it is higher, but during extreme activity it is lower. These oxygen tensions represent the usual limits within which an oxygen carrier must work. An efficient carrier would be nearly fully saturated in the lungs, but should be able to give up most of this to a working muscle.

Let us first see whether a carrier that binds oxygen in a simple equilibrium represented by

$$\text{Oxygen} + \text{carrier} \rightleftharpoons \text{oxygen} \cdot \text{carrier}$$

would be satisfactory. For this type of carrier the dissociation constant would be given by the simple expression

$$K_d = \frac{[\text{oxygen}][\text{carrier}]}{[\text{oxygen} \cdot \text{carrier}]}$$

and the saturation curve would be a rectangular hyperbola. This model would be valid even for a carrier with several oxygen binding

sites per molecule (which we know is the case for hemoglobin) as long as each site were independent and not influenced by the presence or absence of oxygen at adjacent sites.

If such a carrier had a dissociation constant that permitted 90% saturation in the lungs, then, as shown in Figure 23.2A, at a partial pressure of 20 mmHg (2.67 kPa) it would still be 66% saturated, and would have delivered only 24% of what it could carry. This would not be very efficient.

What about some other simple carrier, one that bound oxygen less tightly and therefore released most of it at low partial pressure, so that the carrier was, say, only 20% saturated at 20 mmHg (2.67 kPa)? Again, as shown in Figure 23.2B, it would be relatively inefficient; in the lungs this carrier could fill only 56% of its maximum oxygen capacity, and would deliver only 36% of what it could carry. It appears then that the mere fivefold change in oxygen tension between the lungs and the unloading site is not compatible with efficient operation of a simple carrier. Simple carriers are not sensitive enough to respond massively to a signal as small as a fivefold change.

Figure 23.2 also shows the oxygen binding curve of hemoglobin in normal blood. The curve is sigmoid, not hyperbolic, and it cannot be described by a simple equilibrium expression. Hemoglobin, however, is a very good physiological oxygen carrier. It is 98% saturated in the lungs and only about 33% saturated in the working muscle. Under these conditions it delivers nearly 70% of the oxygen it can carry.

It can be seen in Figure 23.2 that hemoglobin is 50% saturated with oxygen at a partial pressure of 27 mmHg (3.60 kPa). The partial pressure corresponding to 50% saturation is called the P_{50}. P_{50} is the most common way of expressing hemoglobin's oxygen affinity. In analogy with K_m for enzymes, a relatively high P_{50} corresponds to a relatively low oxygen affinity.

It is important to notice that the steep part of hemoglobin's saturation curve lies in the range of oxygen tensions that prevail in the extrapulmonary tissues of the body. This means that relatively small decreases in oxygen tension in these tissues will result in large increases in oxygen delivery. Furthermore, small shifts of the curve to the left or right will also strongly influence oxygen delivery. In Sections 23.4, 23.6, and 23.7 we see how physiological signals effect such shifts and result in enhanced delivery under conditions of increased oxygen demand. Small decreases of oxygen tension in the lungs, however, such as occur at moderately high altitudes, do not seriously compromise hemoglobin's ability to bind oxygen. This will be true as long as the alveolar partial pressure of oxygen remains in a range that corresponds to the relatively flat region of hemoglobin's oxygen dissociation curve.

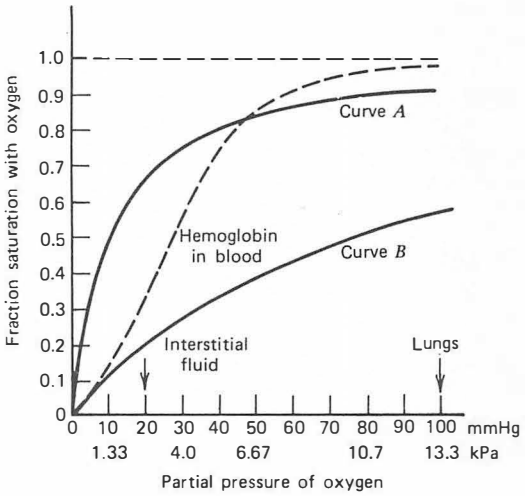


Figure 23.2
Oxygen saturation curves for two hypothetical oxygen carriers and for hemoglobin.
Curve A: Hypothetical carrier with hyperbolic saturation curve (a simple carrier) 90% saturated in the lungs and 66% saturated at the partial pressure found in interstitial fluid. Curve B: Hypothetical carrier with hyperbolic saturation curve (another simple carrier), 56% saturated in the lungs and 20% saturated at the partial pressure found in interstitial fluid. Dotted curve: Hemoglobin in whole blood.

Finally, we can see from Figure 23.2 that the binding of oxygen by hemoglobin is cooperative. At very low oxygen tension the hemoglobin curve tends to follow the hyperbolic curve, which represents relatively weak oxygen binding, but at higher tensions it actually rises above the hyperbolic curve that represents tight binding. Thus it can be said that hemoglobin binds oxygen weakly at low oxygen tension and tightly at high tension. The binding of the first oxygen to each hemoglobin molecule somehow enhances the binding of subsequent oxygens.

23.3 THE HILL EQUATION

Hemoglobin's ability to bind oxygen cooperatively has been known for many years, and for most of that time it stood as a unique example of a system that could not be described by an ordinary chemical equilibrium equation. Early attempts to explain the hemoglobin–oxygen system were not successful, but the early workers were able to write equations that would fit the experimental binding curve. The best known of these is the Hill equation,

$$\frac{Y}{1 - Y} = Kx^n$$

where Y is the fraction of oxygen-binding sites filled with oxygen, K is a constant, x is the partial pressure of oxygen, and n (the Hill coefficient) is a number chosen so that a plot of $\log Y/(1 - Y)$ vs $\log x$ is a straight line. The Hill equation is general, and it describes any chemical system in which one chemical species binds another in a specific manner. If n is given the value of 1, it will describe a simple equilibrium, and for this case the Hill equation can be derived easily from the law of mass action. In cooperative systems the Hill coefficient is greater than 1; in systems that exhibit anticooperative effects it has a value between 0 and 1. The Hill coefficient is thus a measure of the degree of a system's cooperativity. Its maximum value is the number of binding sites per molecule. Hemoglobin, with four oxygen binding sites, could have a maximum Hill coefficient of 4. In fact, hemoglobin's Hill coefficient is about 2.7. This means that its oxygen binding is not as cooperative as might be, but it is certainly more cooperative than would be possible for a molecule with only two cooperating oxygen-binding sites.

23.4 HEMOGLOBIN AND ALLOSTERISM: EFFECT OF 2,3-DIPHOSPHOGLYCERATE

Hemoglobin's binding of oxygen was the original example of a homotropic effect (cooperativity and allosterism are discussed in Chapter 4), but hemoglobin also exhibits a heterotropic effect of great physiological significance. This involves its interaction with 2,3-diphosphoglycerate (DPG). Figure 23.3 shows the structure of DPG; it is closely related to the glycolytic intermediate, 1,3-diphosphoglycerate, from which it is in fact biosynthesized.

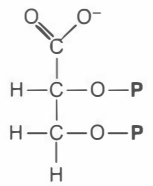


Figure 23.3
2,3-Diphosphoglycerate (DPG).

It had been known for many years that hemoglobin in the red cell bound oxygen less tightly than purified hemoglobin could (Figure 23.4). It had also been known that the red cell contained high levels of DPG, nearly equimolar with hemoglobin. Finally the appropriate experiment was done to demonstrate the relationships between these two facts. It was shown that the addition of DPG to purified hemoglobin produced a shift to the right of its oxygen-binding curve, bringing it into congruence with the curve observed in whole blood. Other organic polyphosphates, such as ATP and inositol hexaphosphate, also have this effect. Inositol hexaphosphate is the physiological

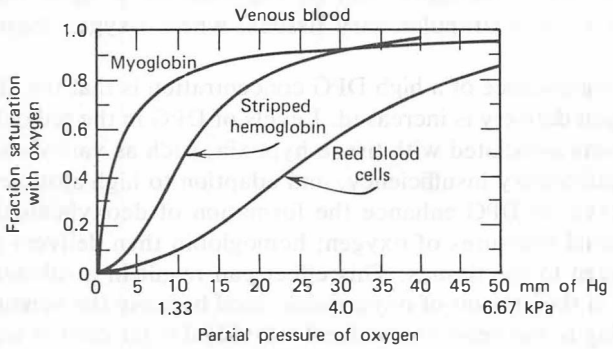


Figure 23.4
Oxygen dissociation curves for myoglobin, for hemoglobin that has been stripped of CO_2 and organic phosphates, and for whole red blood cells.
Data from O. Brenna, et al., *Advan. Exp. Biol. Med.*, 28:19, 1972. Adapted from R. W. McGilvery. *Biochemistry, A Functional Approach*, 2nd ed., W. B. Saunders, Philadelphia, 1979, p. 236.

effector in birds, where it replaces DPG, and ATP plays a similar role in some fish.

Monod's model of allosterism explains heterotropic interaction. Applying this model to hemoglobin, in the deoxyconformation (the T state) a cavity large enough to admit DPG exists between the β chains of hemoglobin. This cavity is lined with positively charged groups, and firmly binds one molecule of the negatively charged DPG. In the oxyconformation (the R state) this cavity is smaller, and it no longer accommodates DPG as easily. The result is that the binding of DPG to oxyhemoglobin is much weaker. Since DPG binds preferentially to the T state, the presence of DPG shifts the R–T equilibrium in favor of the T state; the deoxyhemoglobin conformation is thus stabilized over the oxyhemoglobin conformation.

$$\begin{array}{ccc} \text{Hb(T)} & \rightleftharpoons & \text{Hb(R)} \\ + & & + \\ \text{DPG} & & \text{O}_2 \\ \Updownarrow & & \Updownarrow \\ \text{Hb} \cdot \text{DPG(T)} & & \text{Hb} \cdot \text{O}_2\text{(R)} \end{array}$$

For oxygen to overcome this and bind to hemoglobin, a higher concentration of oxygen is required. Oxygen tension in the lungs is sufficiently high under most conditions to saturate hemoglobin almost completely, even when DPG levels are high. The physiological effect of DPG, therefore, can be expected to be upon release of oxygen in the extrapulmonary tissues, where oxygen tensions are low.

The significance of a high DPG concentration is that the efficiency of oxygen delivery is increased. Levels of DPG in the red cell rise in conditions associated with tissue hypoxia, such as various anemias, cardiopulmonary insufficiency, and adaption to high altitude. These high levels of DPG enhance the formation of deoxyhemoglobin at low partial pressures of oxygen; hemoglobin then delivers more of its oxygen to the tissues. This effect can result in a substantial increase in the amount of oxygen delivered because the venous blood returning to the heart of a normal individual is (at rest) at least 60% saturated with oxygen. Much of this oxygen can dissociate in the peripheral tissues if the DPG concentration rises.

The DPG mechanism works very well as a compensation for tissue hypoxia as long as the partial pressure of oxygen in the lungs remains high enough that oxygen binding in the lungs is not compromised. Since, however, the effect of DPG is to shift the oxygen-binding curve to the right, the mechanism will not compensate for tissue hypoxia when the partial pressure of oxygen in the lungs falls

too low. Then the increased efficiency of oxygen unloading in the tissues is counterbalanced by a decrease in the efficiency of loading in the lungs. This may be a factor in determining the maximum altitude at which people choose to establish permanent dwellings, which is about 18,000 feet (~ 5500m). There is evidence that a better adaption to extremely low ambient partial pressures of oxygen would be a shift of the curve to the left.

23.5 OTHER HEMOGLOBINS

Although hemoglobin A is the major form of hemoglobin in adults and in children over 7 months of age, accounting for about 90% of their total hemoglobin, it is not the only normal hemoglobin species. Normal adults also have 2–3% of hemoglobin A_2, which is composed of two α chains like those in hemoglobin A and two δ chains. It is represented as $\alpha_2\delta_2$. The δ chains are distinct from the β chains, and are under independent genetic control. Hemoglobin A_2 does not appear to be particularly important in normal individuals.

Several species of modified hemoglobin A also occur normally. These are designated A_{1a1}, A_{1a2}, A_{1b}, and A_{1c}. They are adducts of hemoglobin with various sugars, such as glucose, glucose 6-phosphate, and fructose 1,6-bisphosphate. The quantitatively most significant of these is hemoglobin A_{1c}. It arises from the covalent binding of a glucose residue to the N-terminal of the β chain. The reaction is not enzyme-catalyzed, and its rate depends on the concentration of glucose. As a result, hemoglobin A_{1c} forms more rapidly in uncontrolled diabetics and can compose up to 12% of their total hemoglobin under some circumstances. It has been suggested that hemoglobin A_{1c} levels or total glycosylated hemoglobin levels may be a useful measure of how well diabetes has been controlled during the days and weeks before the measurement is taken; measurement of blood glucose only indicates how well it is under control at the time the blood sample is taken.

Fetal hemoglobin, hemoglobin F, is the major hemoglobin component of the newborn. It contains two γ chains in place of the β chains, and is represented as $\alpha_2\gamma_2$. Shortly before birth γ chain synthesis diminishes and β chain synthesis is initiated, and by the age of 7 months well over 90% of the infant's hemoglobin is hemoglobin A.

Hemoglobin F is adapted to the environment of the fetus, who must get his or her oxygen from the maternal blood, a source that is far poorer in oxygen than the atmosphere. In order to compete with

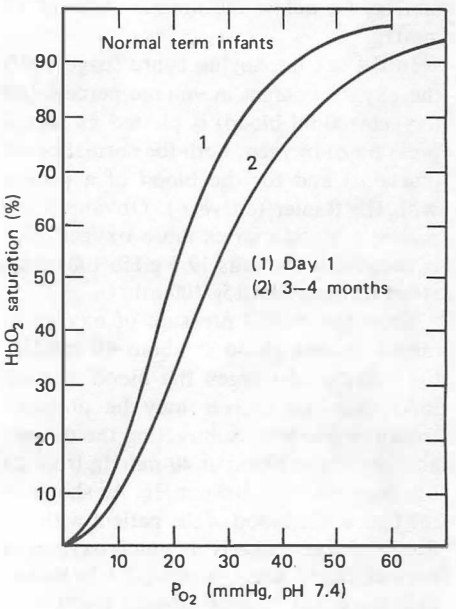


Figure 23.5
Oxygen dissociation curves after birth.
Adapted from F. A. Oski and M. Delivoria-Papadopoulos, *J. Pediatr.*, 77:941, 1970.

the maternal hemoglobin for oxygen, fetal hemoglobin must bind oxygen more tightly; its oxygen binding curve is thus shifted to the left relative to hemoglobin A. This is accomplished not through an intrinsic difference in the oxygen affinities of these hemoglobins but through a difference in the influence of DPG upon them. In hemoglobin F two of the groups that line the DPG-binding cavity have neutral side chains instead of the positively charged ones that occur in hemoglobin A. Consequently, hemoglobin F binds DPG less tightly and thus binds oxygen more tightly than hemoglobin A does. Furthermore, about 15–20% of the hemoglobin F is acetylated at the N-terminals; this is referred to as hemoglobin F_1. Hemoglobin F_1 does not bind DPG, and its affinity for oxygen is not affected at all by DPG. The postnatal change from hemoglobin F to hemoglobin A, combined with a rise in red cell DPG that peaks 3 months after birth, results in a gradual shift to the right of the infant's oxygen binding curve (Figure 23.5). The result is greater delivery of oxygen to the tissues at this age than at birth, in spite of a 30% decrease in the infant's total hemoglobin concentration.

There are many inherited anomalies of hemoglobin synthesis in which there is synthesis of a structurally abnormal hemoglobin; these are called hemoglobinopathies. They may involve the substitution of one amino acid in one of the polypeptide chains for some other amino acid or they may involve deletion of part of the polypeptide chain. In some cases the change is clinically insignificant, but in others it causes serious disease. Most of the known β chain substitutions cause clinical symptoms, whereas most α chain substitutions do not. Evidently the α chain is somehow so important that nearly any alteration of its ability to function normally is incompatible with life. Thus, only the harmless α chain mutations survive. A multitude of functional impairments can occur in the β chain without causing death (Clin. Corr. 23.1).

CLIN. CORR. 23.1 HEMOGLOBINS WITH ABNORMAL OXYGEN AFFINITY

Some abnormal hemoglobins have an altered affinity for oxygen. If oxygen affinity is increased (P_{50} decreased), oxygen delivery to the tissues will be diminished unless some sort of compensation occurs. Typically the body responds by producing more erythrocytes (polycythemia) and more hemoglobin. Hb Ranier is an abnormal hemoglobin in which the P_{50} is 12.9 mmHg, far below the normal value of 27 mmHg.

In the accompanying figure (page 1097) the oxygen content in volume percent (ml oxygen/100 ml blood) is plotted vs partial pressure of oxygen, both for normal blood (curve *a*) and for the blood of a patient with Hb Ranier (curve *b*). Obviously the patient's blood carries more oxygen; this is because it contains 19.5 g Hb/100 ml instead of the usual 15g/100 ml.

Since the partial pressure of oxygen in mixed venous blood is about 40 mmHg, the volume of oxygen the blood of each individual can deliver may be obtained from the graph by subtracting the oxygen content of the blood at 40 mmHg from its oxygen content at 100 mmHg. As shown in the figure, the blood of the patient with Hb Ranier delivers nearly as much oxygen as normal blood does, although Hb Ranier delivers a significantly smaller fraction of the total amount it carries. Evidently polycythemia is an effective compensation for this condition, at least in the resting state.

23.6 ENVIRONMENTAL EFFECTS ON OXYGEN BINDING

Effect of Temperature

Temperature has a significant effect on oxygen binding by hemoglobin, as shown in Figure 23.6. At below-normal temperatures the binding is tighter, resulting in a leftward shift of the curve; at higher temperatures the binding becomes weaker, and the curve is shifted to the right.

The effect of elevated temperature is like that of high levels of DPG, in that both enhance unloading of oxygen. The temperature effect is physiologically useful, as it makes additional oxygen available to support the high metabolic rate found in fever or in exercising muscle with its elevated temperature. The relative insensitivity to temperature of oxygen binding at high partial pressure of oxygen minimizes compromise of oxygen uptake in the lungs under these conditions.

The tighter binding of oxygen that occurs in hypothermic conditions is not consequential in hypothermia induced for surgical purposes. The decreased oxygen utilization by the body and increased solubility of oxygen in plasma at lower temperatures, as well as the increased solubility of carbon dioxide, which acidifies the blood, compensate for hemoglobin's diminished ability to release oxygen.

Effect of pH

Hydrogen ion concentration influences hemoglobin's oxygen binding. As shown in Figure 23.7, low pH shifts the curve to the right, enhancing oxygen delivery, whereas high pH shifts the curve to the left. It is customary to express oxygen binding by hemoglobin as a function of the pH of the plasma because it is this value, not the pH within the erythrocyte, that is usually measured. The pH of the erythrocyte cell sap is lower than the plasma pH, but since these two fluids are in equilibrium, changes in the one reflect changes in the other.

The influence of pH upon oxygen binding is physiologically significant, since a decrease in pH is often associated with increased oxygen demand. An increased metabolic rate results in increased production of carbon dioxide and, as in muscular exercise, lactic acid. Lactic acid is also produced by hypoxic tissue. These acids produced by metabolism help release oxygen to support that metabolism.

The increase in acidity of hemoglobin as it binds oxygen is known as the *Bohr effect;* an equivalent statement is that the Bohr effect is the increase in basicity of hemoglobin as it releases oxygen. The effect may be expressed by the equation

$$HHb + O_2 \rightleftharpoons HbO_2 + H^+$$

Clearly this equation gives the same information as Figure 23.7, namely, that increases in hydrogen ion concentration will favor the formation of free oxygen from oxyhemoglobin, and conversely, that oxygenation of hemoglobin will lower the pH of the solution.

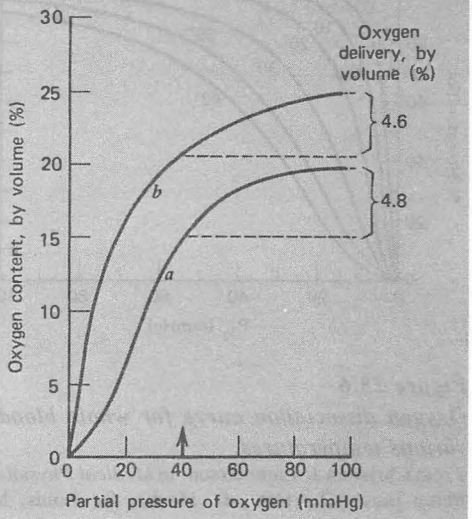


Figure C.C. 23.1

Curve a shows the oxygen dissociation curve of normal blood with a hemoglobin of 15 g/dl, P_{50} 27 mmHg, n 2.8, at pH 7.4, 37°C. Curve b shows that of blood from a patient with Hb Ranier, having a hemoglobin of 19.5 g/dl, P_{50} 12.9 mmHg, n 1.2, at the same pH and temperature. (1 mmHg ≈ 133.3 Pa.) On the right is shown the oxygen delivery. The compensatory polycythemia and hyperbolic curve of Hb Ranier results in practically normal arterial and venous oxygen tensions. Arrow indicates normal mixed venous oxygen tension.

From A. J. Bellingham, Hemoglobins with altered oxygen affinity. *Brit. Med. Bull.*, 32:234, 1976.

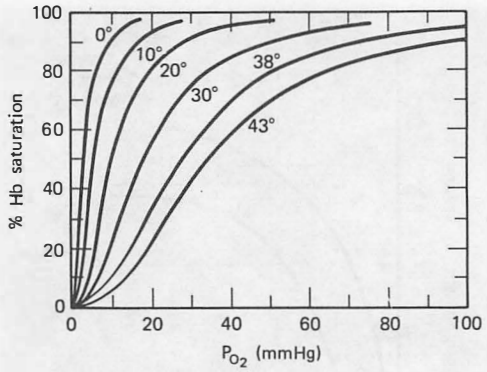


Figure 23.6
Oxygen dissociation curve for whole blood at various temperatures.
From Christian J. Lambertson, in *Medical Physiology*, Philip Bard, ed., 11th ed., Mosby, St. Louis, Mo., 1961, p. 596.

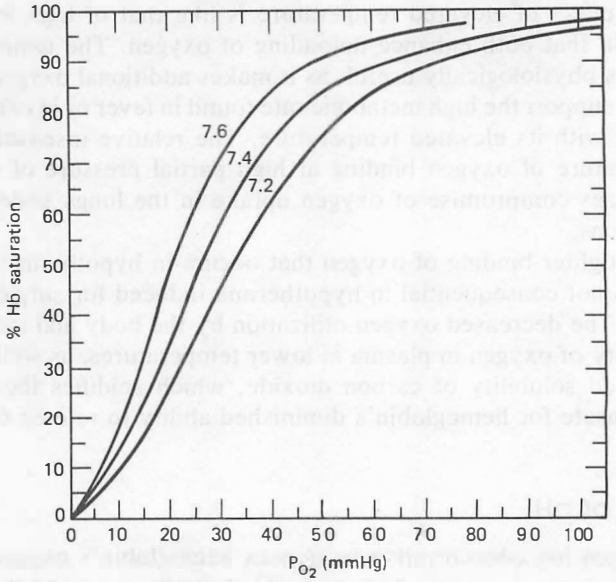


Figure 23.7
Oxygen dissociation curve for whole blood at various values of plasma pH.
Adapted from Christian J. Lambertson, in *Medical Physiology*, Phillip Bard, ed., 11th ed., Mosby, St. Louis, Mo., 1961, p. 596.

23.7 CARBON DIOXIDE TRANSPORT

The carbon dioxide we produce is excreted by the lungs, to which it must be transported by the blood. Carbon dioxide transport is closely tied to hemoglobin and to the problem of maintaining a constant pH in the blood, a problem which will be discussed subsequently.

Carbon dioxide is present in the blood in three major forms, as dissolved CO_2, as HCO_3^- (formed by ionization of H_2CO_3 produced when CO_2 reacts with water), and as carbamino compound (formed when CO_2 reacts with amino groups of protein). Each of these is present both in arterial blood and in venous blood, as shown in the top three lines of Table 23.2. Net transport to the lungs for excretion is represented by the concentration difference between arterial and venous blood, shown in the last column. Notice that for each form of

Table 23.2 Properties of Blood of Man at Rest[a]

	Arterial			*Venous*			Δ		
	Serum	*Cells*	*Blood*	*Serum*	*Cells*	*Blood*	*Serum*	*Cells*	*Blood*
Hb carbamino compound meq/liter blood		*1.13*	*1.13*		*1.42*	*1.42*		*+0.29*	*+0.29*
HCO_3^-, meq/liter blood	13.83	*5.73*	*19.56*	14.84	*6.41*	*21.25*	+1.01	*+0.68*	*+1.69*
Dissolved CO_2, meq/liter blood	0.71	0.48	1.19	0.82	0.56	1.38	+0.11	+0.08	+0.19
Total CO_2, meq/liter blood	14.54	7.34	21.88	15.66	8.39	24.05	+1.12	+1.05	+2.17
Free O_2, mmol/liter blood			0.10			0.04			−0.06
Bound O_2, mmol/liter blood			8.60			6.01			−2.59
Total O_2, mmol/liter blood			8.70			6.05			−2.65
pO_2, mmHg			88.0			37.2			−50.8
pCO_2, mmHg			41.0			47.5			+6.5
pH	7.40	7.19		7.37	7.17		−0.03	−0.02	
Volume, cc/liter blood	551.7	448.3	1000	548.9	451.1	1000	−2.8	+2.8	0.0
H_2O, cc/liter blood	517.5	322.8	840.0	514.7	325.6	840.0	−2.8	+2.8	0.0
Cl^-, meq/liter blood	57.71	24.30	82.01	56.84	25.17	82.01	−0.88	+0.88	0.0

SOURCE: From J. Baggott, The contribution of carbamate to physiological carbon dioxide transport. *Trends in Biochemical Sciences*, 3:N207, 1978, with permission of the publisher.

[a] Hemoglobin, 9 mM; serum protein, 39.8 g/liter of blood; respiratory quotient, 0.82.

carbon dioxide the arterial–venous difference is only a small fraction of the total amount present; venous blood contains only about 10% more total carbon dioxide (total carbon dioxide is the sum of HCO_3^-, dissolved CO_2 and carbamino compound) than arterial blood does.

Carbon dioxide, after it enters the bloodstream for transport, generates hydrogen ions in the blood. Most come from bicarbonate ion formation, which occurs in the following manner.

Carbon dioxide entering the blood diffuses into the erythrocytes. The erythrocyte membrane, like most other biological membranes, is freely permeable to dissolved CO_2. Within the erythrocytes most of the carbon dioxide is acted upon by the intracellular enzyme, carbonic anhydrase, which catalyzes the reaction

$$CO_2 + H_2O \underset{\text{anhydrase}}{\overset{\text{carbonic}}{\rightleftharpoons}} H_2CO_3$$

This reaction will proceed in the absence of a catalyst, as is well known to all who drink carbonated beverages. Without the catalyst, however, it is too slow to meet the body's needs, taking over 100 s to reach equilibrium. Recall that at rest the blood makes a complete circuit of the body in only 60 s. Carbonic anhydrase is a very active enzyme, having a turnover number of the order of 10^6, and inside the

erythrocytes the reaction reaches equilibrium within 1 s, less than the time spent by the blood in the capillary bed. The enzyme is zinc-requiring, and accounts for a portion of our dietary requirement for this metal.

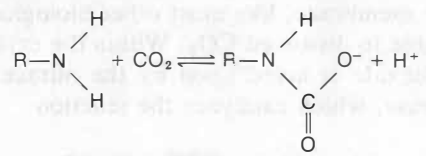
The ionization of carbonic acid

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

is a rapid, spontaneous reaction. It results in the production of equivalent amounts of H^+ and HCO_3^-. Since, as shown in the last column of line 2 in Table 23.2, 1.69 meq of bicarbonate was added to each liter of blood by this process, 1.69 meq of H^+ must also have been generated per liter of blood. The addition of this much acid, over 10^{-3} equiv H^+, to a liter of water would give a final pH below 3. Since the pH of venous plasma has an average value of 7.37, clearly most of the H^+ generated during HCO_3^- production must be consumed by buffer action and/or other processes. This is discussed below.

Because of the compartmentalization of carbonic anhydrase, essentially all of the conversion of CO_2 to carbonic acid, and ultimately to HCO_3^-, occurs inside the erythrocyte. Negligible amounts of CO_2 react nonenzymatically in the plasma. This means that virtually all of the increase in HCO_3^- in venous blood as compared to arterial comes from intraerythrocyte HCO_3^- generation. To be sure, most of this diffuses into the plasma, so that venous plasma HCO_3^- is higher than arterial, but the erythrocyte was the site of its formation.

It has been observed that in the presence of carbonic anhydrase inhibitors, such as acetazolamide or cyanide, blood will still take up a certain amount of carbon dioxide rapidly. This is due to the reaction of carbon dioxide with amino groups of proteins within the erythrocyte to form carbamino compound.



Most carbamino compound forms with the amino groups of hemoglobin. Deoxyhemoglobin forms carbamino compound more readily than oxyhemoglobin does, and oxygenation causes the release of CO_2 that had been bound to deoxyhemoglobin as carbamino compound.

Carbamino compound formation occurs only with uncharged aliphatic amino groups, not with the charged form, $R—NH_3^+$. The

pH within the erythrocyte is normally about 7.2, somewhat more acidic than the plasma. Since amino groups of proteins have pK's well to the alkaline side of 7.2, they will be mostly in the charged (undissociated acid) form. Removal of some of the uncharged form via carbamino compound formation will shift the equilibrium, generating more uncharged amino groups and an equivalent amount of H^+.

$$R - NH_3^+ \rightleftharpoons R - NH_2 + H^+$$

Clearly the formation of carbamino compound is, like HCO_3^- formation, a process that generates H^+. But, while each CO_2 that forms HCO_3^- generates an equivalent amount of H^+, each equivalent of CO_2 that forms carbamino compound can be expected to produce somewhere between 1 and 2 equiv H^+. The amount is somewhat less than 2 because the reestablishment of equilibrium shown in the above equation does not replace *every* R—NH_2 consumed during carbamino compound formation; it merely reestablishes the previous *ratio* of R—NH_2: R—NH_3^+. The lower the pK of the amino group, the closer the acid generation will be to 1 H^+ per CO_2 reacted.

The fact that only uncharged groups can form carbamino compound severely limits the groups that can potentially participate in this reaction. Typical amino groups, such as the ε-amino groups in the side chains of lysyl residues, have pK's ~9.5–10.5. If the pK were 10.2, then at an intracellular pH of 7.2 only one ε-amino group in a thousand would be uncharged and able to react with carbon dioxide. The α-amino groups at the N-terminals of proteins, however, have much lower pK's, in the range of 7.6–8.4. This is because of the electron-withdrawing effect of the nearby oxygen of the peptide linkage. For an amino group with pK = 8.2, 1 out of every 10 molecules would be uncharged inside the cell and able to react with CO_2. A lower pK (or a higher intracellular pH) would result in an even greater availability of the group. Because of their lower pK's the α-amino groups at the N-terminals of hemoglobin's polypeptide chains are the principal sites of carbamino compound formation. If all four N-terminal amino groups of hemoglobin are blocked chemically by reaction with cyanate, carbamino compounds do not form.

The N-terminal amino groups of the β chains form part of the binding site of DPG. Since the N-terminals cannot bind DPG and simultaneously form carbamino compound, a competition arises. CO_2 diminishes the effect of DPG and, conversely, DPG diminishes the ability of hemoglobin to form carbamino compound. Ignorance of the latter interaction led to a major overestimation of the role of carbamino compound in carbon dioxide transport. Prior to the discovery of the DPG effect, careful measurements were made of the

capacity of purified hemoglobin (no DPG present) to form carbamino compound. The results were assumed to be applicable to hemoglobin in the erythrocyte, leading to the erroneous conclusion that carbamino compound accounted for 25–30% or more of carbon dioxide transport. It now appears that 13–15% of carbon dioxide transport is via carbamino compound.

Hemoglobin, in addition to being the primary oxygen carrier and a transporter of carbon dioxide in the covalently bound form of carbamino compound, also plays the major role in handling the hydrogen ions produced in carbon dioxide transport. It does this by buffering and by a second mechanism, which is discussed below. Hemoglobin's buffering power is due to its ionizable groups with pK's in the neighborhood of the intracellular pH of the erythrocyte. These include the four α-amino groups of the N-terminal amino acids and the imidazole side chains of the histidine residues. Hemoglobin has 38 histidines per tetramer; these therefore provide the bulk of hemoglobin's buffering ability.

In whole blood, buffering absorbs about 60% of the acid generated in normal carbon dioxide transport. Although hemoglobin is by far the most important noncarbonic buffer in blood, the organic phosphates in the erythrocytes, the plasma proteins, and so on, also make a significant contribution. Buffering by these compounds accounts for about 10% of the acid, leaving ~ 50% of acid control specifically attributable to buffering by hemoglobin. These buffer systems minimize the change in pH that occurs when acid or base is added, but do not altogether prevent that change. A small difference in pH between arterial and venous blood is therefore observed.

The remainder of the acid arising from carbon dioxide is absorbed by hemoglobin via a mechanism that has nothing to do with buffering. Recall that when hemoglobin became oxygenated it became a stronger acid and released H^+ (the Bohr effect). In the capillaries, where oxygen is released, the opposite occurs:

$$HbO_2 + H^+ \rightleftharpoons HHb + O_2$$

Simultaneously, carbon dioxide enters the capillaries and is hydrated:

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

Addition of these two equations gives

$$HbO_2 + CO_2 + H_2O \rightleftharpoons HHb + HCO_3^- + O_2$$

revealing that to some extent this system can take up H^+ arising from carbon dioxide, and can do so with no change in H^+ concentration

(that is, with no change in pH). Hemoglobin's ability to do this, through the operation of the Bohr effect, is referred to as the *isohydric carriage* of CO_2. As already pointed out, there is a small A-V difference in plasma pH. This is because the isohydric mechanism cannot handle all the acid generated during normal CO_2 transport; if it could, no such difference would occur.

Estimates of the importance of the isohydric mechanism in handling normal respiratory acid production have changed upward and downward over the years. The older, erroneous estimates arose out of a lack of knowledge of the multiple interactions in which hemoglobin participates. The earliest experiments, titrations of purified oxyhemoglobin and purified deoxyhemoglobin, revealed that oxygenation of hemoglobin resulted in release of an average of 0.7 H^+ for every O_2 bound. This figure still appears in textbooks, and much is made of it. Authors point out that with a Bohr effect of this magnitude the isohydric mechanism alone could handle all of the acid produced by the metabolic oxidation of fat (RQ of fat is 0.7), and buffering would be unnecessary. Unfortunately the experimental basis for this interpretation is physiologically unrealistic: the titrations were done in the total absence of carbon dioxide, which we now know binds to some of the Bohr groups, forming carbamino compound and diminishing the effect. When experiments are carried out in the presence of physiological amounts of carbon dioxide, there is a drastic diminution of the Bohr effect, so much so that at pH 7.45 the isohydric mechanism could handle only the amount of acid arising from carbamino compound formation. This work, however, was carried out prior to our appreciation of the competition between DPG and carbon dioxide for the same region of the hemoglobin molecule. Finally, in 1971, careful titrations of whole blood under presumably physiological conditions were carried out, yielding a value of 0.31 H^+ released per O_2 bound. This value is the basis of the present assertion that the isohydric mechanism accounts for about 40% of the acid generated during normal carbon dioxide transport. The quantitative contributions of various mechanisms to the handling of acid arising during carbon dioxide transport are summarized in Table 23.3. The major role of hemoglobin in handling this acid is obvious.

We have seen that essentially all of HCO_3^- formation is intracellular, catalyzed by carbonic anhydrase, and that the vast bulk of the H^+ generated by CO_2 is handled within the erythrocyte. These two observations bear upon the final distribution of HCO_3^- between plasma and the erythrocyte.

Intracellular formation of HCO_3^- increases its intracellular concentration. Since the erythrocyte membrane is freely permeable to HCO_3^- (as well as to certain other ions, including Cl^- and H^+), HCO_3^- will diffuse out of the erythrocyte, increasing the plasma

Table 23.3 Distribution of the H^+ Generated During Normal Carbon Dioxide Transport

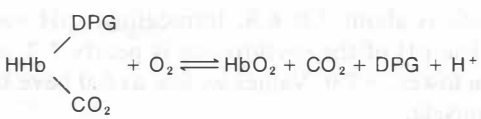
Buffering	
By hemoglobin	50%
By other buffers	10%
Isohydric mechanism (hemoglobin)	40%

HCO_3^- concentration. Electrical neutrality must be maintained across the membrane as this happens. Maintenance of neutrality can be accomplished in principle either by having a positively charged ion accompany HCO_3^- out of the cell or by having some other negatively charged ion enter the cell in exchange for the HCO_3^-. Since the distribution of the major cations, Na^+ and K^+, is under strict control, it is the latter mechanism that is seen and the ion that is exchanged for HCO_3^- is Cl^-. Thus as HCO_3^- is formed in red cells as they pass through the capillary bed, it moves out into the plasma and Cl^- comes in to replace it. The increase in intracellular [Cl^-] is shown in the last line of Table 23.2. In the lungs, where all events of the peripheral tissue capillary beds are reversed, HCO_3^- migrates into the cells to be converted to CO_2 for exhalation, and Cl^- returns to the plasma. The exchange of Cl^- and HCO_3^- between the plasma and the erythrocyte is called the *chloride shift*.

The intracellular buffering of H^+ from carbon dioxide causes the cells to swell, giving venous blood a slightly (0.6%) higher hematocrit than arterial blood. (The hematocrit is the volume percent of red cells in the blood.) This occurs because the charge on the hemoglobin molecule becomes more positive with every H^+ that binds to it. Each bound positive charge requires an accompanying negative charge to maintain neutrality. Thus as a result of buffering there is a net accumulation of HCO_3^- or Cl^- inside the erythrocyte. An increase in the osmotic pressure of the intracellular fluid results from this increase in concentration of particles. As a consequence water migrates into the cells, causing them to swell slightly. Typically, an arterial hematocrit might be 44.8 and a venous hematocrit, 45.1, as shown in Table 23.2 by the line labeled "volume, cc/liter blood."

23.8 INTERRELATIONSHIPS AMONG HEMOGLOBIN, OXYGEN, CARBON DIOXIDE, HYDROGEN, AND 2,3-DIPHOSPHOGLYCERATE

By now it should be clear that multiple interrelationships of physiological significance exist among the ligands of hemoglobin. These interrelationships may be summarized schematically as follows:



This equation shows that changes in the concentration of H^+, DPG, or CO_2 have similar effects on oxygen binding. The equation will help you remember the effect of changes in any one of these variables upon hemoglobin's oxygen affinity.

DPG levels in the red cell are controlled by product inhibition of its synthesis and by pH. Hypoxia results in increased levels of deoxyhemoglobin on a time-averaged basis. Since deoxyhemoglobin binds DPG more tightly, in hypoxia there is less free DPG to inhibit its own synthesis, and so DPG levels will rise due to increased synthesis. The effect of pH is that high pH increases DPG synthesis and low pH decreases DPG synthesis. Since changes in DPG levels take many hours to become complete, this means that the immediate effect of a decrease in blood pH is to enhance oxygen delivery by the Bohr effect. If the acidosis is sustained (most causes of chronic metabolic acidosis are not associated with a need for enhanced oxygen delivery), diminished DPG synthesis leads to a decrease in intracellular DPG concentration, and hemoglobin's oxygen affinity returns toward normal. Thus we have a system that can respond appropriately to acute conditions, such as vigorous exercise, but which when faced with a prolonged abnormality of pH readjusts to restore normal (and presumably optimal) oxygen delivery.

23.9 INTRODUCTION TO pH REGULATION

When we considered carbon dioxide transport we noted the large amount of H^+ generated by this process, and we considered the ways in which the blood pH was kept under control. Control of blood pH is important because changes in blood pH will cause changes in intracellular pH, which in turn may profoundly alter metabolism. Protein conformation is affected by pH, as is enzyme activity. In addition, the equilibria of important reactions that consume or generate hydrogen ions, such as any of the oxidation–reduction reactions involving pyridine nucleotides, will be shifted by changes in pH.

The normal arterial plasma pH is 7.40 ± 0.05; the pH range com-

patible with life is about 7.8–6.8. Intracellular pH varies with the type of cell. The pH of the erythrocyte is nearly 7.2, whereas most other cells are lower, ~7.0. Values as low as 6.0 have been reported for skeletal muscle.

It is fortunate for both diagnosis and treatment of diseases that the acid–base status of the intracellular fluid influences and is influenced by the acid–base status of the blood. Blood is readily available for analysis, and when alteration of body pH becomes necessary, intravenous administration of acidifying or alkalizing agents is efficacious.

23.10 BUFFER SYSTEMS OF PLASMA, INTERSTITIAL FLUID, AND CELLS

Every differentially permeable membrane in the body separates the solution that it encloses from the other body water compartments. Each type of compartment contains characteristic kinds and concentrations of solutes, some of which are buffers at physiological pH's. Although the solutes in the cytoplasm of each type of cell are different, most cells are similar enough that they are considered together for purposes of acid–base balance. Thus there are, from this point of view, three major body water compartments: plasma, which is contained in the circulatory system; interstitial fluid, the fluid that bathes the cells; and intracellular fluid.

The compositions of these fluids are given in Figure 23.8. In plasma the major cation is Na^+; small amounts of K^+, Ca^{2+}, and Mg^{2+} are also present. The two dominant anions are HCO_3^- and Cl^-. Smaller amounts of protein, phosphate, and SO_4^{2-} are also found, along with a mixture of organic anions (amino acids, etc.), each of which would be insignificant if taken separately. The sum of the anions equals, of course, the sum of the cations. It is apparent at a glance that the composition of interstitial fluid is very similar. The major difference is that interstitial fluid contains much less protein than plasma contains (the capillaries are not normally permeable to the plasma proteins) and, correspondingly, a lower cation concentration. Plasma and interstitial fluid taken together are called the extracellular fluid, and low molecular weight components equilibrate fairly rapidly between the two. H^+, for example, equilibrates between the plasma and ISF within about half an hour. The composition of intracellular fluid is strikingly different. K^+ is the major ca-

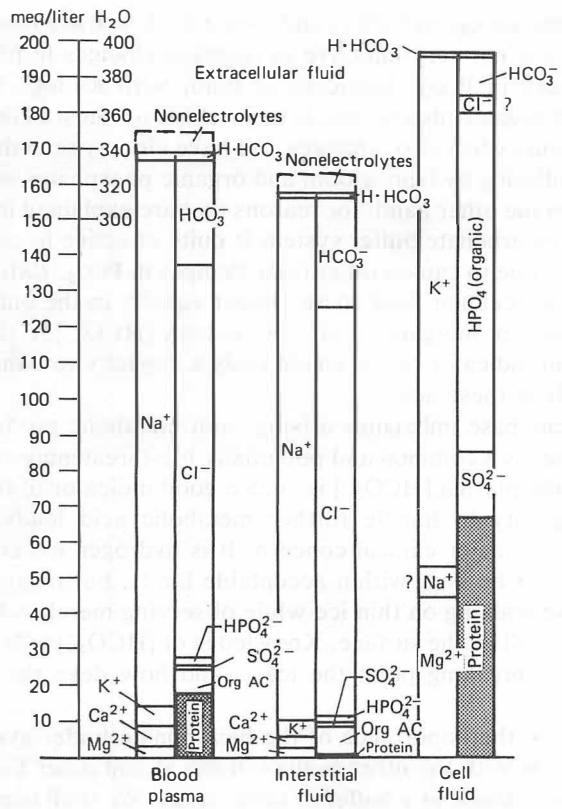


Figure 23.8
Diagram showing chief chemical constituents of the three fluid compartments.
Height of left half of each column indicates total concentration of cations; that of right half, concentration of anions. Both are expressed in meq/liter of water. Note that chloride and sodium values in cell fluid are questioned. It is probable that at least in muscle, intracellular phase contains some sodium but no chloride.
Modified from Gamble. From Magnus I. Gregersen, in *Medical Physiology*, Philip Bard, ed., 11th ed., Mosby, St. Louis, Mo., 1961, p. 307.

tion, while organic phosphates (ATP, DPG, glycolytic intermediates, etc.) and protein are the major anions.

As a result of the differences among these fluid compartments, each fluid makes a different contribution to buffering. The major buffer of the extracellular fluid, for example, is the HCO_3^-/CO_2 system. Since the pK of the HCO_3^-/CO_2 system is 6.1 (Table 23.4 lists

Table 23.4 Acid Dissociation Constants of Major Physiological Buffers

Buffer System	pK
HCO_3^-/CO_2	6.1
Phosphate	
$HPO_4^{2-}/H_2PO_4^-$	6.7–7.2
Organic phosphate esters	6.5–7.6
Protein	
Histidine side chains	5.6–7.0
N-terminal amino groups	7.6–8.4

the major physiological buffers and their pK's), extracellular fluid at a pH of 7.4 is not very effective in resisting changes in pH arising from changes in P_{CO_2}. Intracellular fluid, with its high levels of protein and organic phosphates, is responsible for most of the buffering that occurs when P_{CO_2} changes. We have already seen the importance of buffering by hemoglobin and organic phosphates within the red cell. On the other hand, for reasons that are explained in Section 23.11, the bicarbonate buffer system is quite effective in controlling pH changes due to causes other than changes in P_{CO_2}. Extracellular fluid and intracellular fluid share almost equally in the buffering of strong organic or inorganic acids. The plasma $[HCO_3^-]$ is, therefore, an excellent indicator of the whole body's capacity to handle additional loads of these acids.

Since acid–base imbalance arising from metabolic production of organic acids is a common and potentially life-threatening condition, and since the plasma $[HCO_3^-]$ is such a good indicator of the whole body's capacity to handle further metabolic acid loads, plasma $[HCO_3^-]$ is of major clinical concern. It is hydrogen ion concentration that must be kept within acceptable limits, but measuring pH alone is like walking on thin ice while observing merely whether or not you are still on the surface. Knowledge of $[HCO_3^-]$ tells you how close to the breaking point the ice is and how deep the water is underneath.

Because of the importance of the bicarbonate buffer system and its interaction with the other buffers of blood and other tissues, we shall consider blood as a buffer in some detail. We shall begin with a brief consideration of a model buffer.

Every buffer consists of a weak acid, HA, and its conjugate base, A. Examples of conjugate base/weak acid pairs include acetate$^-$/acetic acid, NH_3/NH_4^+ and $HPO_4^{2-}/H_2PO_4^-$. Notice that the weak acid may be neutral, positively charged or negatively charged, and that its conjugate base must (since a H^+ been lost) have one less positive charge (or one more negative charge) than the weak acid.

The degree of ionization of a weak acid depends on the concentration of free hydrogen ions. This may be expressed in the form of the Henderson–Hasselbalch equation (derived on page 13) as follows:

$$\text{pH} = \text{p}K + \log \frac{[\text{conjugate base}]}{[\text{acid}]}$$

This is a mathematical rearrangement of the fundamental equilibrium equation. It states that there is a direct relationship between the pH and the ratio of conjugate base : acid. It is important to realize that this *ratio,* not the absolute concentration of any particular species, is the factor that is related to pH. Use of this equation will

help you to understand the action of and to predict the effects of various alterations upon acid–base balance in the body.

Blood plasma is a mixed buffer system; in the plasma the major buffers are HCO_3^-/CO_2, $HPO_4^{2-}/H_2PO_4^-$ and protein/HProtein. The pH is the same throughout the plasma, so each of these buffer pairs distributes independently according to its own Henderson–Hasselbalch equation

$$\begin{aligned} \text{pH} &= \text{p}K_1 + \log \frac{[HCO_3^-]}{[CO_2]} \\ &= \text{p}K_2 + \log \frac{[HPO_4^{2-}]}{[H_2PO_4^-]} \\ &= \text{p}K_3 + \log \frac{[\text{protein}^-]}{[\text{HProtein}]} \end{aligned}$$

Because each pK is different the [conjugate base]/[acid] ratio is also different for each buffer pair. Notice, though, if the ratio is known for any given buffer pair, one automatically has information about the others (assuming the pK's are known).

23.11 THE CARBON DIOXIDE–BICARBONATE BUFFER SYSTEM

As we have seen, the major buffer of the plasma (and of the interstitial fluid as well) is the bicarbonate buffer system. The bicarbonate system has two peculiar properties that make its operation unlike that of typical buffers. We shall examine this important buffer in some detail, since a firm understanding of it is the key to a grasp of acid–base balance.

In the first place, the component which we consider to be the acid in this buffer system is CO_2, which is not truly an acid, but an acid anhydride. It reacts with water to form carbonic acid, which is indeed a typical weak acid.

$$CO_2 + H_2O \rightleftharpoons H_2CO_3$$

Carbonic acid then rapidly ionizes to give H^+ and HCO_3^-.

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

If these two equations are added, H_2CO_3 cancels, and the sum is

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

Elimination of H_2CO_3 from formal consideration is realistic since, not only does it simplify matters, but H_2CO_3 is, in fact, quantitatively insignificant. Because the equilibrium of the reaction,

$$CO_2 + H_2O \rightleftharpoons H_2CO_3$$

lies far to the left, H_2CO_3 is present only to the extent of 1/200 of the concentration of dissolved CO_2. Since the concentration of water is virtually constant, it need not be included in the equilibrium expression for the reaction, and one may write:

$$K = \frac{[H^+]\,[HCO_3^-]}{[CO_2]}$$

The value of K is 7.95×10^{-7}.

The concentration of a gas in a solution is proportional to the partial pressure of the gas. Thus we measure the partial pressure of CO_2 (P_{CO_2}). P_{CO_2} is then multiplied by a conversion factor, α, to get the millimolar concentration of dissolved CO_2.

$$\alpha P_{CO_2} = \text{mM}$$

α has a value of 0.03 mmol/liter mmHg (or 0.225 mmol/liter kPa) at 37°C. The equilibrium expression thus becomes

$$K = \frac{[H^+][HCO_3^-]}{0.03 \cdot P_{CO_2}}$$

and the Henderson–Hasselbalch equation for this buffer system becomes (after taking the negative logarithm of K)

$$pH = 6.1 + \log \frac{[HCO_3^-]}{0.03 \cdot P_{CO_2}}$$

$[HCO_3^-]$ is expressed in units of meq/liter.

We said earlier that the bicarbonate buffer system, with a pK of 6.1, was not effective against carbonic acid in the pH range of 7.8–6.8, but that it was effective against noncarbonic acids. The usual rules of chemical equilibrium dictate that a buffer is not very useful in a pH range more than about one unit beyond its pK. Thus what needs to be explained is how the bicarbonate system can be effective against

noncarbonic acids: its failure to buffer carbonic acid is expected. The manner in which it buffers noncarbonic acids in a pH range far from its pK is the second unusual property of this buffer system. Notice that the explanation of this property in the following paragraph involves the flow of materials in a living system, and so departs from mere equilibrium considerations.

Consider first a typical buffer, consisting of a mixture of a weak acid and its conjugate base. When a strong acid is added, most of the added H^+ combines with the conjugate base. As a result, [weak acid] increases and simultaneously [conjugate base] diminishes. The ratio of [conjugate base]/[weak acid] therefore changes, and so does the pH. Of course, the pH changes much less than if there were no buffer present. Now imagine a system in which the weak acid, as it is generated by the reaction of the added strong acid with the conjugate base, is somehow removed so that while [conjugate base] diminishes, [weak acid] remains nearly constant. In this case the ratio of [conjugate base]/[weak acid] would change much less for a given addition of strong acid, and the pH would also change much less. This is exactly what happens with the bicarbonate buffer system in the body. As strong acid is added, $[HCO_3^-]$ diminishes, and CO_2 is formed. But the excess CO_2 is exhaled, so that the ratio of $[HCO_3^-]/\alpha P_{CO_2}$ does not change so dramatically. In like manner, if strong base is added to the body, it will be neutralized by carbonic acid, but CO_2 will be replaced by metabolism, and again, the ratio of $[HCO_3^-]/\alpha P_{CO_2}$ will not change as much as would be expected. The bicarbonate buffer system in the body is thus an *open* system in which the P_{CO_2} term is adjusted to meet the body's needs. If respiration should be unable to accomplish this adjustment, then P_{CO_2} would change strikingly, and the bicarbonate system would be relatively ineffective, in keeping with the prediction of chemical equilibrium.

A graphical representation of the Henderson–Hasselbalch equation for the bicarbonate buffer system is a valuable aid to learning and understanding how this system reflects the acid–base status of the body. One of the most common of these representations is the pH–bicarbonate diagram, shown in Figure 23.9. A $[HCO_3^-]$ of up to 40 meq/liter is shown on the ordinate; this is adequate to deal with most situations. Similarly, since plasma pH does not exceed 7.8 or (except transiently) fall below 7.0 in living patients, the abscissa of the graph is limited to the range of 7.0–7.8. The normal plasma $[HCO_3^-]$, 24 meq/liter, and the normal plasma pH, 7.4, are indicated. The third variable, CO_2, can be shown on a two-dimensional graph by assigning a fixed value to P_{CO_2} and then showing, *for that value*, the relationship between pH and $[HCO_3^-]$. Figure 23.9 shows that relationship when P_{CO_2} has its normal physiological value of 40 mmHg

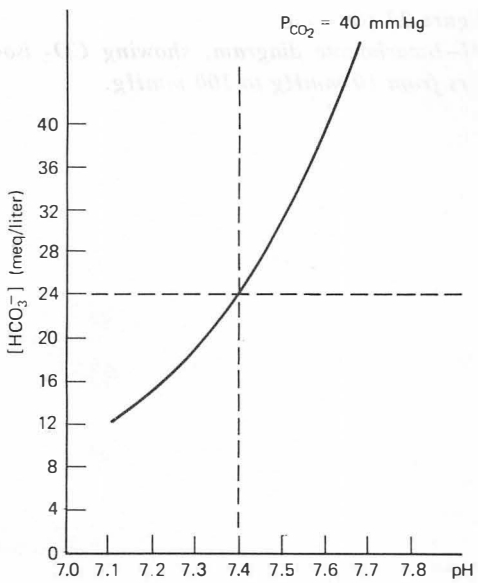


Figure 23.9
pH–Bicarbonate diagram, including the 40 mmHG (5.33 kPa) CO_2 isobar, and showing the normal values of plasma pH and bicarbonate ion concentration.

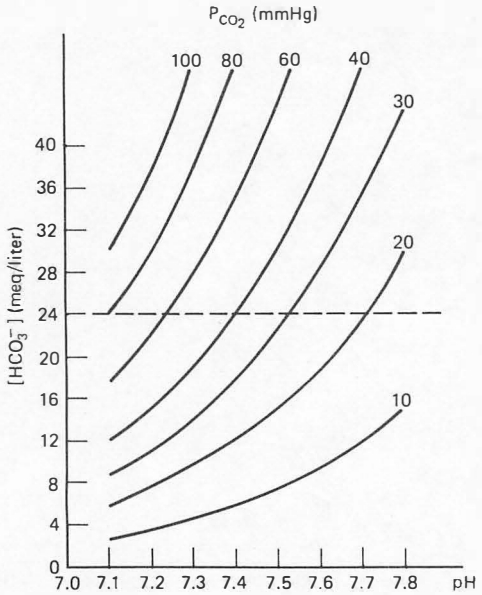


Figure 23.10
pH–Bicarbonate diagram, showing CO_2 isobars from 10 mmHg to 100 mmHg.

(5.33 kPa). The line is called the 40 mmHg (5.33 kPa) isobar. Whenever P_{CO_2} is 40 mmHg (5.33 kPa), pH and $[HCO_3^-]$ must be somewhere on that line.

In a similar manner we can plot isobars for various abnormal values of P_{CO_2}. These curves are shown in Figure 23.10. The range of values given covers those found in patients. Any point on the graph gives the values of the three variables of the Henderson–Hasselbalch equation for the bicarbonate system at that point. Since all you need to locate a point are any two of these variables, the third can be read directly from the graph.

Let us now see how the bicarbonate buffer system behaves when it is in the presence of other buffers, as it is in whole blood. First, let us acidify the system by increasing the concentration of the acid-producing component, CO_2. For every CO_2 that reacts with water to produce a H^+, one HCO_3^- will also form. Most of the H^+, however, will be buffered by protein and phosphate. As a result, $[HCO_3^-]$ will rise much more than $[H^+]$. Similarly, if acid is removed from this system by decreasing P_{CO_2}, $[HCO_3^-]$ will decrease. $[H^+]$ will not decrease by an equivalent amount, though, because the other buffers will dissociate to resist the pH change. The results of these processes as they occur in whole blood, with its various intracellular and extracellular buffers, are shown in Figure 23.11. Let us start at the point that represents the normal values: pH of 7.4, $[HCO_3^-]$ of 24 meq/liter and P_{CO_2} of 40 mmHg (5.33 kPa). As P_{CO_2} rises to 80 mmHg (10.7 kPa), bicarbonate goes up to 28 meq/liter, an increase of 4 meq/liter. This means H_2CO_3 must have increased by 4 meq/liter, and that it immediately ionized, giving H^+ and HCO_3^-. The pH, however, drops to 7.18; this represents an increase in $[H^+]$ of only 26×10^{-6} meq/liter. The other 3.999974 meq/liter of H^+ produced by the ionization of carbonic acid were taken up by the phosphate, hemoglobin, plasma protein, and other buffer systems. If P_{CO_2} were to decrease, the opposite would occur. Thus by altering P_{CO_2} in the presence of HCO_3^- and other buffers, a line is generated with a definite nonzero slope. For the blood system, this is called the buffering line of blood. Notice that if P_{CO_2} is the only variable that is changed, the response of the system is confined to movements along this line.

The slope of the buffering line depends on the concentration of the nonbicarbonate buffers. If they were more concentrated, they would better resist changes in pH. An increase in P_{CO_2} to 80 mmHg (10.7 kPa) would then cause a smaller drop in pH, and since the more concentrated buffers would react with more hydrogen ions (produced by the ionization of carbonic acid), $[HCO_3^-]$ would rise higher. Thus the slope of the buffering line would be steeper.

Hemoglobin is quantitatively the second most important blood buffer, exceeded only by the bicarbonate buffer system. Since

hemoglobin concentration in the blood can fluctuate widely in various disease states, it is the most important physiological determinant of the slope of the blood buffer line. Figure 23.12 shows how the slope of the blood buffer line varies with hemoglobin concentration.

Having now seen how the bicarbonate buffer system in blood responds to changes in P_{CO_2} and how this response is modified by changing the hemoglobin concentration, let us examine the response of blood to the addition of noncarbonic acids such as HCl, acetoacetic acid, and so on. We shall continue to analyze the situation in terms of the pH–bicarbonate diagram. The starting point will again be the normal state: pH = 7.4, $[HCO_3^-]$ = 24 meq/liter, and P_{CO_2} = 40 mmHg (5.33 kPa). As acid is added, it will react with all the blood buffers, and the concentrations of all of the conjugate bases will decrease. Since the bicarbonate system is the major blood buffer, the decrease in $[HCO_3^-]$ will be substantial. If P_{CO_2} is held constant at 40 mmHg (5.33 kPa) as a noncarbonic acid is added, the changes in the system can be represented by a point sliding down the 40 mmHg (5.33 kPa) isobar, as shown in Figure 23.13. If alkali is added to the blood, all the undissociated acids of the various buffer systems will participate in neutralizing it. Again, if this occurs at a fixed P_{CO_2} of 40 mmHg (5.33 kPa), the changes in the system will be represented by a point sliding up the 40 mmHg (5.33 kPa) isobar.

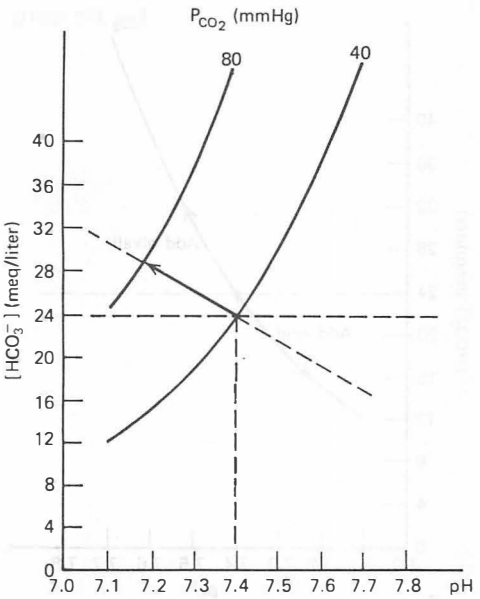


Figure 23.11
The buffering line of blood.
This pH–bicarbonate diagram shows the changes in pH and $[HCO_3^-]$ that occur in whole blood in vitro when P_{CO_2} is changed. Notice that the relationship between pH and $[HCO_3^-]$ is described by a straight line with a nonzero slope.

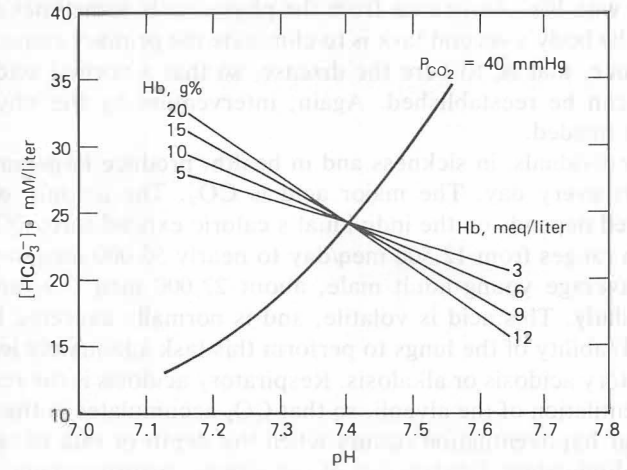


Figure 23.12
Slope of the buffering line of blood as it various with hemoglobin concentration.
From H. W. Davenport, *The ABC of Acid–Base Chemistry*, 6th ed. revised, University of Chicago Press, Chicago, Ill., 1974, p. 55.

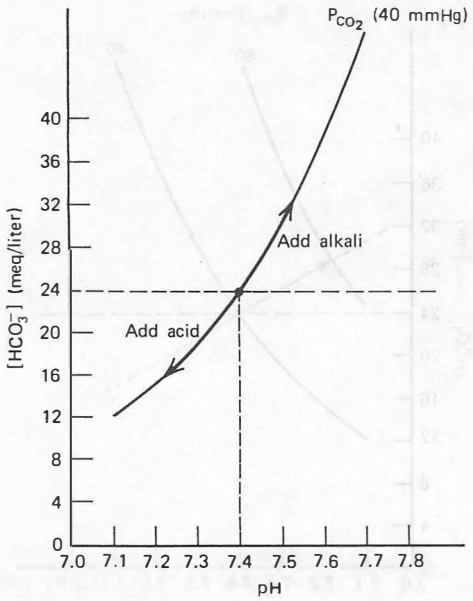


Figure 23.13
The effect of adding noncarbonic acid or alkali to whole blood with P_{CO_2} fixed at 40 mmHg.

Notice that, just as changes in P_{CO_2} were represented by points confined to the blood buffer line, changes due to the addition of acid or base at a fixed P_{CO_2} are represented by points confined to the CO_2 isobar.

The effects on blood of changing P_{CO_2} or of adding acid or alkali, as we have just described, are realistic qualitative models of what can happen in certain disease states. In the following sections we shall see how these changes occur in the body and how the body compensates for them.

23.12 ACID–BASE BALANCE AND ITS MAINTENANCE

It should come as no surprise that mechanisms exist whereby the body normally rids itself of excess acid or alkali. The physiological implication is that if a patient is in a state of continuing acidosis (excess acid or deficiency of alkali in the body) or alkalosis (excess alkali or deficiency of acid in the body), there must be a continuing cause of the imbalance. In such a situation the body's first task is to somehow compensate so plasma pH does not exceed the limits compatible with life. Assistance from the physician is sometimes necessary. The body's second task is to eliminate the primary cause of the imbalance, that is, to cure the disease, so that a normal acid–base status can be reestablished. Again, intervention by the physician may be needed.

All individuals, in sickness and in health, produce large amounts of acids every day. The major acid is CO_2. The amount of CO_2 produced depends on the individual's caloric expenditure; CO_2 production ranges from 12,500 meq/day to nearly 50,000 meq/day, and in an average young adult male, about 22,000 meq CO_2 are produced daily. This acid is volatile, and is normally excreted by the lungs. Inability of the lungs to perform this task adequately leads to respiratory acidosis or alkalosis. Respiratory acidosis is the result of hypoventilation of the alveoli, so that CO_2 accumulates in the body. Alveolar hypoventilation occurs when the depth or rate of respiration is diminished. Obstruction of the airway, neuromuscular disorders and diseases of the central nervous system are common causes of acute respiratory acidosis. Chronic respiratory acidosis is seen in patients with chronic obstructive lung disease, such as emphysema. Obviously, since the common element in all these conditions is increased alveolar P_{CO_2}, inhalation of a gas mixture with a high P_{CO_2} could also cause respiratory acidosis.

Respiratory alkalosis, on the other hand, arises from decreased alveolar P_{CO_2}. Hyperventilation due to anxiety is probably the most common cause. Central nervous system injury involving the respiratory center, salicylate poisoning, fever, and artificial ventilation are other causes. At high altitude, due to the decrease in total atmospheric pressure, alveolar P_{CO_2} also falls, producing chronic respiratory alkalosis.

Various amounts of nonvolatile acids are also produced by the body. The diet and the physiological state of the individual determine the kinds and amounts of these acids. Oxidation of sulfur-containing amino acids produces H^+ and SO_4^{2-}, the equivalent of sulfuric acid. Hydrolysis of phosphate esters is equivalent to the formation of phosphoric acid. The contribution of these processes depends on the amount of acid precursors ingested; for an individual consuming an average American diet there is a net daily acid production of about 60 meq.

Metabolism normally makes certain amounts of lactic acid, acetoacetic acid and β-hydroxybutyric acid. In some physiological and pathological states these are produced in excess, and accumulation of the excess causes acidosis. When an ammonium salt of a strong acid, such as ammonium chloride, or when arginine hydrochloride or lysine hydrochloride is administered, it is converted to urea, and the corresponding strong acid (hydrochloric acid in these examples) is synthesized. Ingestion of salicylates, methyl alcohol, or ethylene glycol results in production of strong organic acids. Accumulation of any of these nonvolatile acids leads to metabolic acidosis.

While it is obvious that excess acid production can cause acidosis, the same net effect can arise from abnormal loss of base, as could be predicted from the Henderson–Hasselbalch equation for the bicarbonate buffer system. Renal tubular acidosis is a condition in which this occurs. Abnormal amounts of HCO_3^- from the blood escape into the urine, leaving the body acidotic. A more common cause of bicarbonate depletion is severe diarrhea. In this chapter it will be assumed that kidney function is normal.

Mammals do not synthesize alkaline compounds from neutral starting materials. Metabolic alkalosis therefore arises from intake of excess alkali or abnormal loss of acid. An alkali commonly taken by many people is sodium bicarbonate. A less obvious source of alkali is the salt of any metabolizable organic acid. Sodium lactate is often administered to combat acidosis; normal metabolism converts it to sodium bicarbonate. The net reaction is as follows:

$$Na^- + CH_3CHOHCOO^- + 3O_2 \rightleftharpoons$$

$$Na^+ + HCO_3^- + 2CO_2 + 2H_2O$$

Most fruits and vegetables have a net alkalizing effect on the body for this reason. They contain a mixture of organic acids (which are metabolized to CO_2 and H_2O, and therefore have no long term effect on acid–base balance), and salts of organic acids, which give rise to bicarbonate. Abnormal loss of acid, as can occur with prolonged vomiting or gastric lavage, causes alkalosis. Alkalosis may also be produced by rapid loss of body water, as in diuresis, which may temporarily increase $[HCO_3^-]$ in the plasma and extracellular fluid.

Role of the Kidney

Excess nonvolatile acid and excess bicarbonate are excreted by the kidney. As a result the pH of the urine varies as a function of the body's need to excrete these materials. For an individual on a typical American diet urine pH is~6, indicating a net acidification as compared to plasma. This is consistent with our knowledge that the typical diet results in a net production of acid. The pH of the urine can range from a lower limit of 4.4 up to 8.0.

A typical daily urine volume is about 1.2 liters. At the minimum urine pH of 4.4, $[H^+]$ is only 4×10^{-2} meq/liter, and it would take 1,250 liters of urine to excrete 50 meq of acid as free hydrogen ions. Clearly most of the acid we excrete must be in some form other than H^+. A form that can be excreted in a reasonable concentration, such as $H_2PO_4^-$ or NH_4^+, is needed.

Let us now see how the kidney accomplishes the excretion of acid or base. Figure 23.14 shows the fundamental functioning unit of the kidney, the nephron. Each human kidney contains several million of these. They serve first to filter the blood and then to modify the filtrate into urine.

Filtration occurs in the glomerulus, which consists of a tuft of capillaries enclosed by an epithelial envelope called the glomerular capsule (formerly Bowman's capsule). Water and low molecular weight solutes, such as inorganic ions, urea, sugars, and amino acids (but not normally substances with molecular weights above 70,000, such as plasma proteins), escape from these capillaries and collect in the capsular space. This ultrafiltrate of plasma then passes through the proximal convoluted tubule, where most of the water and solutes are reabsorbed. The tubule fluid continues through the loop of the nephron (loop of Henle) and through the distal convoluted tubule, where further reabsorption of some solutes and secretion of others occurs. The tubule fluid then passes into the collecting tubule, where concentration can occur if necessary. The fluid may now be called urine; it contains 1% or less of the water and solutes of the original glomerular filtrate.

The kidney regulates acid–base balance by controlling bicarbon-

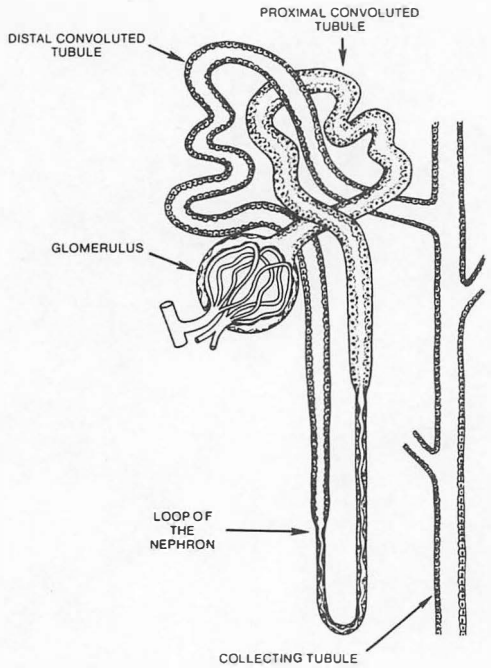


Figure 23.14
Diagram showing the essential features of a typical nephron in the human kidney.
Adapted from Homer W. Smith, *The Physiology of the Kidney*, Oxford University Press, London, 1937, p. 6.

ate reabsorption and by secreting acid. Both of these processes depend on formation of H^+ and HCO_3^- from CO_2 and water within the tubule cells, shown in Figure 23.15A. The H^+ formed in this reaction is then actively secreted into the tubule fluid in exchange for a Na^+. Sodium uptake by the tubule cell is partly passive, with Na^+ flowing down the electrochemical gradient, and partly active, via a Na^+–H^+ antiport system. At this point sodium has been reabsorbed in exchange for a H^+, and sodium bicarbonate has been generated within the tubule cell. The sodium bicarbonate is then pumped out of the cell into the interstitial fluid, which equilibrates with the plasma.

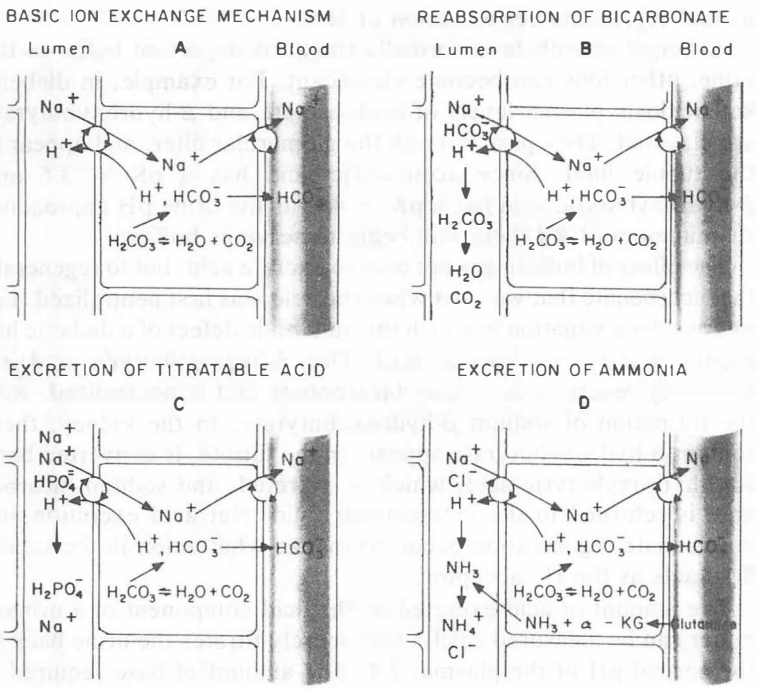


Figure 23.15
Role of the exchange of tubular cell H^+ ions in tubular fluid in renal regulation of acid–base balance.
(*A*) *Basic ion exchange mechanism.* (*B*) *Reabsorption of bicarbonate.* (*C*) *Excretion of titratable acid.* (*D*) *Excretion of ammonia.*
Adapted from R. E. Pitts, Role of ammonia production and excretion in regulation of acid–base balance, *New Engl. J. Med.*, 284:32, 1971, with permission of the publisher.

The H^+ which has been secreted into the tubule fluid can now experience one of three fates. First, it can react with a HCO_3^-, as shown in Figure 23.15B, to form CO_2 and water. The overall net effect of this process is to move sodium bicarbonate from the tubule fluid back into the interstitial fluid. The name given to this is reabsorption of sodium bicarbonate.

As reabsorption of sodium bicarbonate proceeds, the tubule fluid becomes depleted of HCO_3^-, and the pH drops from its initial value, which was identical to the pH of the plasma from which it was derived. As HCO_3^- becomes less available and the pH comes closer to the pK of the $HPO_4^{2-}/H_2PO_4^-$ buffer system, more and more of the H^+ will be taken up by this buffer. Buffering is the second fate the H^+ can experience, and it is represented in Figure 23.15C. $H_2PO_4^-$ is not readily reabsorbed by the kidney. It passes out in the urine, and its loss represents net excretion of H^+.

Although phosphate is normally the most important buffer in the urine, other ions can become significant. For example, in diabetic ketoacidosis plasma levels of acetoacetate and β-hydroxybutyrate are elevated. They pass through the glomerular filter, and appear in the tubule fluid. Since acetoacetic acid has a pK = 3.6 and β-hydroxybutyric acid has a pK = 4.7, as the urine pH approaches its minimum of 4.4 these will begin to serve as buffers.

The effect of buffering is not only to excrete acid, but to regenerate the bicarbonate that was lost when the acid was first neutralized. Let us consider a situation in which the metabolic defect of a diabetic has produced β-hydroxybutyric acid. This β-hydroxybutyric acid immediately reacts with sodium bicarbonate and is neutralized, with the formation of sodium β-hydroxybutyrate. In the kidney, then, sodium β-hydroxybutyrate appears in the filtrate, is converted back to β-hydroxybutyric acid, which is excreted, and sodium bicarbonate is returned to the extracellular fluid. Net acid excretion and bicarbonate regeneration occur no matter what anion in the tubule fluid acts as the H^+ acceptor.

The amount of acid excreted as the acid component of a urinary buffer can be measured easily. One merely titrates the urine back to the normal pH of the plasma, 7.4. The amount of base required is identical to the amount of acid excreted in this form, and is referred to as the *titratable acidity* of the urine.

The formation of titratable acidity accounts for about one-third to one-half of our normal daily acid excretion. It is thus an important mechanism for acid excretion, and is capable of putting out as much as 250 meq of acid daily. There is, however, a limit to the amount of acid that can be excreted in this manner. Titratable acidity can be increased only by lowering the pH of the urine or by increasing the

concentration of buffer in the urine, and neither of these processes can proceed indefinitely. The urine pH cannot go below about 4.4; evidently the sodium-for-hydrogen exchange mechanism is incapable of pumping H^+ out of the tubule cells against more than a thousandfold concentration gradient. Buffer excretion is limited not only by the solubility of the buffer, but by limitations to the supply of the buffer ion and of the cations that are necessarily part of the important buffer systems. If, for example, a 600 meq/day acid load were excreted as NaH_2PO_4, the body would be totally depleted of sodium in less than a week.

The third fate the H^+ can experience in the tubule fluid is neutralization by NH_3. The tubule cells can deamidate glutamine, forming glutamate and NH_3, as shown in Figure 23.15D. At a normal intracellular pH about 1% of the ammonia in the cell will be in the uncharged form. This form diffuses rapidly through the cell membrane and appears in the tubule lumen, where it then reacts with the H^+, and forms NH_4^+. NH_4^+ cannot easily diffuse through the cell membrane. Therefore, since the tubule fluid is more acid than the intracellular fluid, extraction of NH_3 from the cell into the tubule fluid occurs. Elimination of NH_4^+ in the urine contributes to net acid excretion.

NH_4^+ is normally a major urinary acid. Typically, one-half to two-thirds of our daily acid load is excreted as NH_4^+. For three reasons it becomes even more important in acidosis. In the first place, since the pK of NH_4^+ is 9.3, acid can be excreted in this form without lowering the pH of the urine, whereas formation of titratable acidity requires a decrease in urine pH. Second, enormous amounts of acid can be excreted in this form. Ammonia is readily available from amino acids, and in prolonged acidosis the NH_4^+ excretion system becomes activated. This activation, however, takes several days; it does not begin to adapt until after 2–3 days, and the process is not complete until 5–6 days after the onset of acidosis. Once complete, though, amounts of acid in excess of 500 meq can be excreted daily as NH_4^+. The third role of NH_4^+ in acidosis is that it spares the body's stores of Na^+ and K^+. Excretion of titratable acid, such as $H_2PO_4^-$, and of the anions of strong acids, such as acetoacetate, require simultaneous excretion of a cation to maintain electrical neutrality. At the onset of acidosis this role is filled by Na^+, and as the body's Na^+ stores become depleted, K^+ excretion rises. If NH_4^+ did not then become available even a moderate acidosis could quickly become fatal.

Total acid excretion, the *total acidity* of the urine, is the sum of the titratable acidity and NH_4^+. Strictly speaking, one should subtract from this sum the urinary HCO_3^-, but this correction is seldom made

in practice. Obviously, in severe metabolic acidosis, where the total acid excretion would be of greatest interest, the urine would be so acidic that HCO_3^- would be nil.

In alkalosis the role of the kidney is simply to allow HCO_3^- to escape. Metabolic alkalosis is therefore seldom long-lasting unless alkali is continuously administered or HCO_3^- elimination is somehow prevented. HCO_3^- elimination may be restricted if the kidney receives a strong signal to conserve Na^+ at a time when there is a deficiency of an easily reabsorbable anion, such as Cl^-, to be reabsorbed with it. Some diuretics cause this. The first renal response is to put out K^+ in exchange for Na^+ from the tubule fluid, and when K^+ stores are depleted, H^+ is exchanged for Na^+. This results in the production of an acidic urine by an alkalotic patient. If NaCl is administered, alkalosis associated with volume and Cl^- depletion may correct itself.

23.13 COMPENSATORY MECHANISMS

We have defined four primary types of acid–base imbalances and we have seen their chemical causes. Respiratory acidosis arises from an increased plasma P_{CO_2}. Respiratory alkalosis is caused by a decreased plasma P_{CO_2}. In metabolic acidosis addition of strong organic or inorganic acid (or loss of HCO_3^-) results in a decreased plasma $[HCO_3^-]$. Conversely, in metabolic alkalosis loss of acid from the body or ingestion of alkali raises the plasma $[HCO_3^-]$. Recall that in an acute respiratory acid–base imbalance, as long as the body has not attempted to compensate, the pH will be abnormal, and the $[HCO_3^-]$ will be somewhere on the buffer line. In an acute metabolic acid–base imbalance, if the patient has made no attempt to compensate, the pH will be abnormal and the $[HCO_3^-]$ will be somewhere on the 40mm isobar.

When the plasma pH deviates from the normal range, various compensatory mechanisms begin to operate. The general principle of compensation is that, since an abnormal condition has directly altered one of the terms of the $[HCO_3^-]/[CO_2]$ ratio, the plasma pH can be readjusted back toward normal by a compensatory alteration of the other term. For example, if a diabetic becomes acidotic due to excess production of ketone bodies, plasma $[HCO_3^-]$ will decrease. Compensation would involve decreasing the plasma $[CO_2]$ so that the $[HCO_3^-]/[CO_2]$ ratio, and therefore the pH, is readjusted back

toward normal. Notice that compensation does not involve a return of $[HCO_3^-]$ and $[CO_2]$ toward normal. Rather, compensation is a secondary alteration in one of these, an alteration that has the effect of counteracting the primary alteration in the other. The result is that the plasma pH is readjusted toward normal. That this is necessarily so is evident from the Henderson–Hasselbalch equation.

$$pH = 6.1 + \log \frac{[HCO_3^-]}{0.03\ P_{CO_2}}$$

If $[HCO_3^-]$ changes, the only way to restore the original $[HCO_3^-]/[CO_2]$ ratio is to change P_{CO_2} in the *same direction*. If the primary change is in P_{CO_2}, the original ratio can be restored only by altering $[HCO_3^-]$ in the *same direction*.

Although some compensatory mechanisms begin to operate rapidly and produce their effects rapidly, others are slower. Several stages of compensation may therefore be seen. First, is the acute stage, before any significant degree of compensation could possibly occur. After the acid–base imbalance has been in effect for a period of time the patient may become *compensated*. This means the compensatory mechanisms have come into play in a normal manner, as expected on the basis of experience with other individuals with an acid–base imbalance of similar type and degree. The "compensated state" does not necessarily imply that the plasma pH is within the normal range. Alternatively, the patient may show no sign of compensation, even though compensation is expected. This state is referred to as *uncompensated;* it arises because compensation cannot occur due to some other abnormality. Finally, there is an intermediate state where compensation is occurring but is not yet as complete as it should be. This is the *partially compensated* state. Factors which limit the compensatory processes will be discussed at the end of this section.

Let us now follow the course of acute onset of each type of acid–base imbalance and the compensatory process. Each of these will be schematically illustrated in a pH–bicarbonate diagram.

Imagine an individual in normal acid–base balance who goes into an acute respiratory acidosis as a result of breathing a gas mixture containing a high level of CO_2. As P_{CO_2} rises, plasma pH will drop and $[HCO_3^-]$ will rise. (If a decrease in pH and a simultaneous rise in $[HCO_3^-]$ suddenly seems anomalous, turn back to Figure 23.11 and the text on page 1112, and review the blood buffer line.) The point describing his or her condition will follow the buffer line to point A, as shown in Figure 23.16. Eventually a new steady-state P_{CO_2} will be established in the alveoli and in the blood, and no further change in P_{CO_2} will occur. The abnormal condition has fixed this patient on an

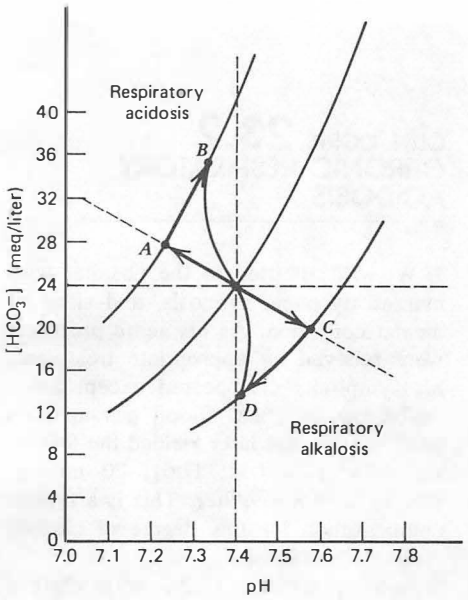


Figure 23.16
pH–Bicarbonate diagram showing compensation for respiratory acidosis (normal state to point B) and for respiratory alkalosis (normal state to point D).

CLIN. CORR. 23.2 CHRONIC RESPIRATORY ACIDOSIS

H.W. was admitted to the hospital with marked dyspnea, cyanosis, and signs of mental confusion. As his acute problems were relieved by appropriate treatment, his symptoms disappeared except for a continuing dyspnea. Blood gas analysis performed 8 days later yielded the following data: pH, 7.32; P_{CO_2}, 70 mmHg; $[HCO_3^-]$, 34.9 meq/liter. This is a typical compensation for this degree of chronic respiratory acidosis.

Another patient, C.Q., with chronic obstructive lung disease was found to have arterial plasma pH, 7.40; $[HCO_3^-]$, 35.9 meq/liter; and P_{CO_2}, 60 mmHg. A return of the plasma pH to 7.4 is not expected in the average patient with respiratory acidosis. Some patients are capable of attaining it, although it becomes less and less likely as P_{CO_2} increases. In this case, with a P_{CO_2} of 60 mmHg, a plasma pH of 7.4 lies outside the 95% probability range. Close questioning of the patient revealed that he had surreptitiously been taking a relative's thiazide diuretic, which superimposed a metabolic alkalosis upon respiratory acidosis.

abnormal high CO_2 isobar. If the condition is returned to normal, he can drop back to the 40 mmHg (5.33 kPa) isobar and all will be well, but until that time all compensatory processes are confined to the higher CO_2 isobar. Compensation will, of course, consist of renal excretion of H^+. Since this is a bicarbonate-producing process, $[HCO_3^-]$ should rise, even though it is already above normal. This could have been predicted from the pH–HCO_3^- diagram with no knowledge of the renal mechanism of compensation: Since it is assumed that the individual is fixed on the high CO_2 isobar by the abnormal condition, the only way the pH can possibly be adjusted toward normal is by sliding up the isobar to point *B* in Figure 23.16. This movement is necessarily linked to an increase in $[HCO_3^-]$. Thus the correct analysis of this compensation could be made either from an understanding of the nature of the compensatory mechanism or from an appreciation of the physical chemistry of the bicarbonate buffer system as expressed in the pH–HCO_3^- diagram. (See Clin. Corr. 23.2.)

Although the path we have described, up the buffer line to point *A* and then up the isobar to point *B*, is a real possibility, it is likely that a respiratory acidosis would develop gradually, with compensation occurring simultaneously. The points describing this progress would fall on the curved line from the normal state to point *B*.

In sudden onset respiratory alkalosis P_{CO_2} drops rapidly. The pH rises and $[HCO_3^-]$ falls, following the buffer line to point *C* in Figure 23.16. As with respiratory acidosis, unless the cause of the decreased alveolar P_{CO_2} is removed, the patient is fixed on an abnormal CO_2 isobar. (See Clin. Corr. 23.3.) Compensation consists of renal excretion of HCO_3^-; plasma $[HCO_3^-]$ diminishes (at a fixed, subnormal P_{CO_2}), and the plasma pH decreases toward normal. This is described in Figure 23.16 by movement along the isobar from point *C* to point *D*. With a gradual onset of respiratory alkalosis, the bicarbonate buffer system would follow points along the curved line from the normal state to point *D*.

In metabolic acidosis two mechanisms are usually available for dealing with the excess acid. The kidneys increase their H^+ excretion, but this takes time, and is not adequate to return $[HCO_3^-]$ and the pH to normal. The other mechanism, which begins to operate almost instantly, is respiratory compensation. The acidosis stimulates the respiratory system to hyperventilate, decreasing the P_{CO_2}. Thus, if onset of a primary metabolic acidosis is represented in Figure 23.17 by a fall in plasma $[HCO_3^-]$ along the 40 mmHg (5.33 kPa) isobar from the normal state to point *E*, the compensatory decrease in P_{CO_2}, and the concomitant rise in pH will be along the line from *E* to *F*. Notice that this line is parallel to the buffer line, and so compensation for a metabolic acidosis involves not only the expected

decrease in P_{CO_2}, but also a *further* small decrease in $[HCO_3^-]$. This is due to the same factor that causes the buffer line itself to have a slope: titration of the nonbicarbonate buffers. The inevitability and the magnitude of the further decrease in $[HCO_3^-]$ can be seen clearly in the pH–bicarbonate diagram.

The principles governing compensation for metabolic alkalosis are like those for metabolic acidosis, but everything is operating in the opposite direction. In metabolic alkalosis the primary defect is an increase in plasma $[HCO_3^-]$; it rises from the normal state to point *G* in Figure 23.17. The immediate physiological response is hypoventilation, followed by increased renal excretion of HCO_3^-. As a result of the hypoventilation the P_{CO_2} increases along the line from *G* to *H*, and a further small rise in $[HCO_3^-]$ occurs.

The respiratory response to a metabolic acid–base imbalance is rapid; an acute metabolic imbalance will not generally be seen outside the experimental laboratory. Indeed, if a physician sees a patient whose plasma pH, $[HCO_3^-]$, and P_{CO_2} are consistent with an acute metabolic imbalance, he concludes that the patient's compensatory mechanisms are impaired and that the patient cannot compensate. The patient would be suffering a mixed respiratory and metabolic acidosis or a mixed respiratory and metabolic alkalosis. Obviously, if a patient had a primary acidosis of one type (respiratory or metabolic) and a primary alkalosis of the other, both caused by independent diseases, the effects of the two on plasma pH would tend to cancel. But even if the pH were within the normal range due to such a circumstance, $[HCO_3^-]$ and P_{CO_2} would be abnormal.

How complete can the process of compensation be? Can the body totally compensate (bring the pH back to the normal range) for any imbalance? Generally, the answer is no. The organs used in compensation, the lungs and the kidneys, do not exist exclusively to deal with acid–base imbalance. There is a limit to how much one can hyperventilate; it is simply impossible to move air into and out of the lungs at an indefinitely high rate for an indefinitely long time. Also, one cannot suspend respiration merely to raise P_{CO_2} to some desired level. The kidney, too, has limits. As the P_{CO_2} rises above 70 mmHg (9.33 kPa) in respiratory acidosis, renal mechanisms for reabsorbing HCO_3^- fail to keep pace, and further increases in plasma $[HCO_3^-]$ are only about what could be expected from titration of the nonbicarbonate buffers. In respiratory alkalosis renal excretion of excess HCO_3^- can, with time, be sufficient to return the plasma pH to within the normal range. Individuals who dwell at high altitude are typically in compensated respiratory alkalosis with their plasma pH within the normal range. For the other types of acid–base imbalance the exact degree of compensation expected of a patient with a given clinical picture is well worked out, but a detailed discussion is be-

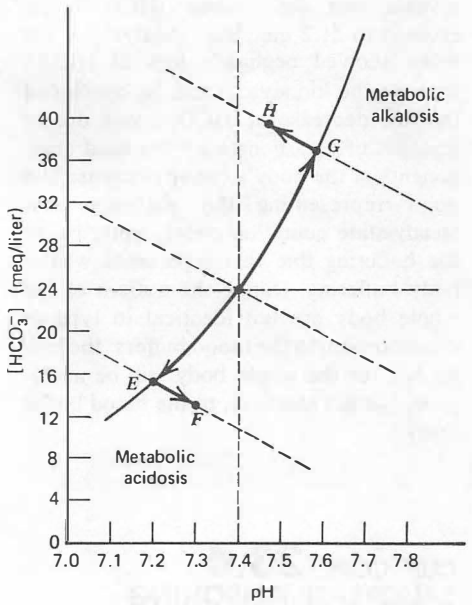


Figure 23.17
pH–Bicarbonate diagram showing compensation for metabolic acidosis (normal state to point F) and for metabolic alkalosis (normal state to point H).

CLIN. CORR. 23.3 ACUTE RESPIRATORY ALKALOSIS

An anesthetized surgical patient with a uretheral catheter in place was hyperventilated as an adjunct to the general anesthesia. Prior to hyperventilation normal values of plasma P_{CO_2} and pH were obtained. Alveolar ventilation was then increased mechanically, and a new steady state was reached, in which the plasma P_{CO_2} was 25 mmHg, and the pH was 7.55. Plasma HCO_3^- was not directly measured, but interpolation from a pH–bicarbonate diagram (e.g., Figure 23.10) or calculation from the Henderson–Hasselbalch equation reveals that the plasma $[HCO_3^-]$ decreased to 21.2 meq/liter. Analysis of the urine showed negligible loss of HCO_3^- through the kidneys. It can be concluded that the decrease in $[HCO_3^-]$ was due to titration of bicarbonate by the acid components of the body's buffer systems. The point representing the patient's new steady-state condition clearly must be on the buffering line that represents whole-body buffering. (Since the buffers of the whole body are not identical in type or concentration to the blood buffers, the buffer line for the whole body will be analogous, but not identical, to the blood buffer line.)

CLIN. CORR. 23.4 SALICYLATE POISONING

Salicylates are the most common cause of poisoning in children. A typical pathway of salicylate intoxication is plotted in the accompanying figure. The first effect of

yond the scope of this chapter. Suffice it to say that if a patient is compensating, but not as well as expected, this is taken to mean that the patient cannot compensate appropriately and must therefore have a mixed acid–base disturbance.

23.14 ALTERNATIVE MEASURES OF ACID–BASE IMBALANCE

Modern clinical laboratories generally report plasma bicarbonate concentration, and the value is used by the physician just as we have used it here. Some laboratories however, report total plasma CO_2. Total plasma CO_2 as reported by the clinical laboratory is the sum of bicarbonate and dissolved CO_2, and so is always slightly higher than $[HCO_3^-]$. At pH 7.4, for example, the ratio of $[HCO_3^-]$ to CO_2 is 20 : 1 (dissolved CO_2 is only 1/21 of the total CO_2); if $[HCO_3^-]$ is 24 meq/liter, $[CO_2]$ is 1.2 meq/liter and total CO_2 is 25.2 meq/liter. At pH 7.1, HCO_3^- is still 10 times as concentrated as dissolved CO_2. Because the major contributor to total CO_2 is HCO_3^-, total CO_2 is often used in the same manner as bicarbonate to make clinical judgments. Strictly speaking, total CO_2 also includes carbamino compound, but current clinical laboratory practice is to ignore carbamino compound when making a blood gas and pH report. If, however, carbamino compound were included in a total CO_2 measurement it would not change the interpretation of the measurement, since carbamino compound, like dissolved CO_2, is only a small fraction of the total CO_2. (See Clin. Corr. 23.4.)

The clinical importance of bicarbonate as a gauge of the whole body's ability to buffer further loads of metabolic acid has given rise to several ways of expressing what the $[HCO_3^-]$ would be if there were no respiratory component or respiratory compensation involved in a patient's condition. The *base excess* is one of the more common of these expressions. It is defined as the amount of acid that would have to be added to the blood to titrate it to pH 7.4 at a P_{CO_2} of 40 mmHg (5.33 kPa) at 37°C. Since the titration is carried out at the normal P_{CO_2}, only the metabolic contribution to acid–base imbalance (primary metabolic imbalance *and* nonrespiratory compensatory processes) would be measured. If a blood sample were acidic under the conditions of the titration, alkali would have to be added instead of acid, and the base excess would be negative.

The concept and the quantitation of base excess are most easily understood if we refer to the pH–bicarbonate diagram. In our dis-

cussion of the blood buffer line we saw how increasing the P_{CO_2} in blood, where other buffers are present, would result in a rise in $[HCO_3^-]$ and a virtually identical decrease in the concentration of other buffer bases. This was because equivalent amounts of the other buffer bases were consumed as they buffered carbonic acid. Since virtually all the carbonic acid formed was buffered, for every HCO_3^- formed one conjugate base of some other system was consumed. In this situation the *total* base in the blood is not measurably changed; only the distribution of HCO_3^- and nonbicarbonate buffer conjugate base is changed. Thus, as long as one remains on the blood buffer line, $[HCO_3^-]$ can change but total base will not. There will be no positive or negative base excess.

If, however, renal activity, diet or some metabolic process adds or removes HCO_3^-, then a positive or negative base excess will be seen. The patient's status will no longer be described by a point on the buffer line, and the base excess will be the difference between the observed plasma $[HCO_3^-]$ and the $[HCO_3^-]$ on the buffer line at the same pH. This is shown in Figure 23.18. In order to calculate this difference, the position of the buffer line (which can be determined from knowledge of the slope and the point representing the normal state) must be known. In the clinical laboratory it can be estimated by measuring the hemoglobin concentration and assuming that it is the major nonbicarbonate buffer.

The buffer line, then, is the dividing line between positive and negative base excess. Any point above it is in the region of positive base excess, and any point below it is in the region of negative base excess. This gives rise to situations that may seem peculiar at first. In Figure 23.19 the $[HCO_3^-]$ at point *A* is normal, but the patient has a negative base excess. A positive or negative base excess occurs as a result of *compensation* for a respiratory acid–base imbalance or *directly* from a metabolic one. Respiratory compensation for a metabolic acid–base imbalance, since it involves movement along a line parallel to the buffer line (Figure 23.17), would cause no further change in the value of the base excess. Clin. Corr. 23.5 involves consideration of base excess.

23.15 THE SIGNIFICANCE OF Na^+ AND Cl^- IN ACID–BASE IMBALANCE

An important concept in diagnosing certain acid–base disorders is the so-called anion gap. Most clinical laboratories routinely measure

salicylate overdose is stimulation of the respiratory center, resulting in respiratory alkalosis. Renal compensation occurs, lowering the plasma $[HCO_3^-]$. A second, delayed effect of salicylate may then appear, metabolic acidosis. Since $[HCO_3^-]$ had been lowered by the previous compensatory process the victim is at a particular disadvantage in dealing with the metabolic acidosis. In addition, but not shown in the graph, respiratory stimulation sometimes persists after the acidosis has run its course. Rational management of salicylate intoxication requires knowledge of the plasma pH and the plasma $[HCO_3^-]$ or its equivalent throughout the course of the condition.

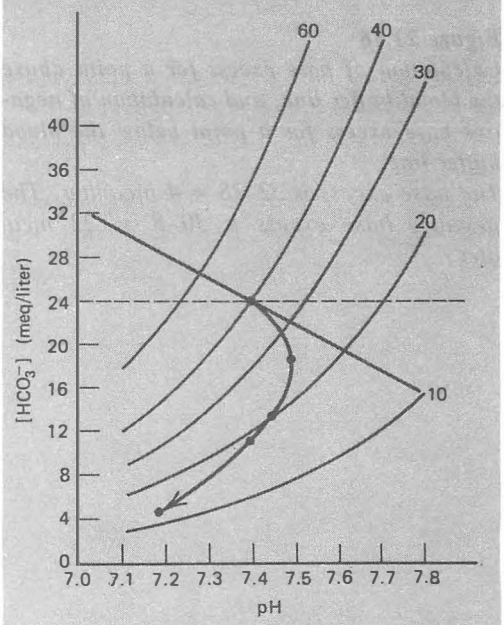


Figure C.C. 23.4
Data replotted from R. B. Singer, The acid–base disturbance in salicylate intoxication. Medicine, 33:1–13, 1954.

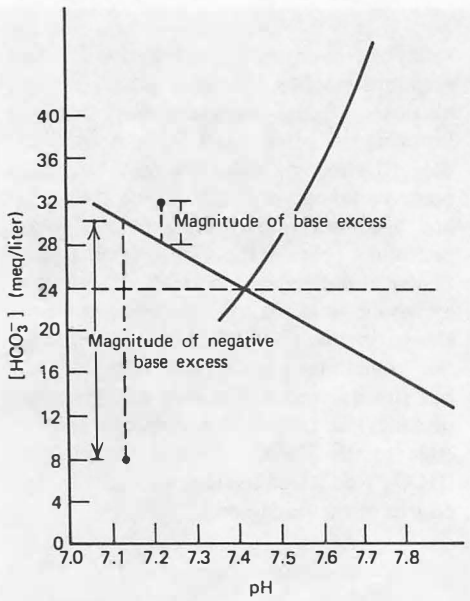


Figure 23.18
Calculation of base excess for a point above the blood buffer line, and calculation of negative base excess for a point below the blood buffer line.
The base excess is 32–28 = 4 meq/liter. The negative base excess is 30–8 = 22 meq/liter.

plasma Na^+, K^+, Cl^-, and HCO_3^-. A glance back at the graph in Figure 23.8 will confirm that in the plasma of a normal individual the sum of Na^+ and K^+ is greater than the sum of Cl^- and HCO_3^-. This difference is called *A*, the anion gap; it represents the other plasma anions (Figure 23.8), which are not routinely measured. It is calculated as follows:

$$A = (Na^+ + K^+) - (Cl^- + HCO_3^-)$$

The normal value of *A* is in the range of 12–16 meq/liter. In some clinical laboratories K^+ is not measured; then the normal value is 8–12 meq/liter. The gap is changed only by conditions that change the sum of the cations or the sum of the anions, or by conditions that change both sums by different amounts. Thus administration or depletion of sodium bicarbonate would not change the anion gap because $[Na^+]$ and $[HCO_3^-]$ would be affected equally. Metabolic acidosis due to HCl or NH_4Cl administration would also leave the anion gap unaffected; here $[HCO_3^-]$ would decrease, but $[Cl^-]$ would increase by an equivalent amount, and the sum of $[HCO_3^-]$ plus $[Cl^-]$ would be unchanged. In contrast, diabetic ketoacidosis or methanol poisoning involves production of strong organic acids which react with HCO_3^-, decreasing its concentration. But since the $[HCO_3^-]$ is replaced by some organic anion, the sum of $[HCO_3^-]$ plus $[Cl^-]$ decreases, and the anion gap increases.

The anion gap is most commonly used to establish a differential diagnosis for metabolic acidosis. In a metabolic acidosis with an increased anion gap, H^+ must have been added to the body with some anion other than chloride. Metabolic acidosis without an increased anion gap must be due either to accumulation of H^+ with chloride or to a decrease in the concentration of sodium bicarbonate. Thus, on the basis of the anion gap, certain diseases can be ruled out, while certain others would have to be considered. This information can be especially important in dealing with patients who cannot give good histories.

The electrolytes of the body fluids interact with each other in a multitude of ways. One of the most important of these involves the capacity for K^+ and H^+ to substitute for one another under certain circumstances. This can occur in the cell, where as we have seen, K^+ is the major cation. In acidosis intracellular $[H^+]$ rises, and it replaces some of the intracellular K^+. The displaced K^+ appears in the plasma, and in time is excreted by the kidneys. This leaves the patient with normal plasma $[K^+]$ (normokalemia), but with seriously depleted body K^+ stores (hypokalia). Subsequent excessively rapid correction of the acidosis may then reverse events. As the plasma pH rises, K^+ flows back into the cells, and plasma $[K^+]$ may decline

to the point where muscular weakness sets in, and respiratory insufficiency may become life-threatening.

In the kidney the reciprocal relationship between K^+ and H^+ results in an association between metabolic alkalosis and hypokalemia. If hypokalemia arises from long-term insufficiency of dietary potassium or long-term diuretic therapy, K^+ levels in the cells will diminish, and intracellular $[H^+]$ will rise. This leads to increased acid excretion, acidic urine and an alkaline arterial plasma pH. We have already seen how in an alkalotic individual a hormonal signal to absorb Na^+ can lead to K^+ loss and then to an exacerbation of the metabolic alkalosis (page 1120). The association also operates in the opposite direction, with alkalosis leading to hypokalemia. In this case increased amounts of $Na^+ + HCO_3^-$ are presented to the distal convoluted tubule, where all K^+ secretion normally takes place (all *filtered* K^+ is reabsorbed; K^+ loss is due to distal tubular secretion). The distal tubule takes up some of the Na^+, but since HCO_3^- does not readily follow across that membrane, the increased Na^+ uptake is linked to increased K^+ secretion. K^+ excretion is complicated, with its control under the influence of a variety of hormones and other factors. The end result, however, is that metabolic alkalosis and hypokalemia go hand in hand, so much so that in some circles the term, "hypokalemic alkalosis," is used synonymously with metabolic alkalosis (Clin. Corr. 23.6).

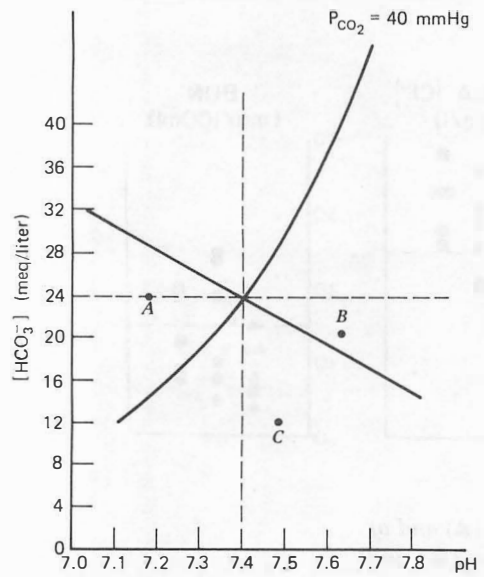


Figure 23.19
Examples showing the sign of the base excess at various points.
At points A and C there is a negative base excess. At point B the base excess is positive.

CLIN. CORR. 23.5 EVALUATION OF CLINICAL ACID–BASE DATA

In a 1972 study of total parenteral nutrition of infants, it was found that infants who received amino acids in the form of a hydrolyzate of the protein fibrin maintained normal acid–base balance. In contrast, infants receiving two different mixtures of synthetic amino acids, FreAmine and Neoaminosol, became acidotic. Both synthetic mixtures contained adequate amounts of all the essential amino acids, but neither contained aspartate or glutamate. The fibrin hydrolyzate contained all of the common amino acids.

The accompanying figure shows the blood acid–base data from these infants. Notice that the normal values for infants, given by the dotted lines, are not quite the same as normal values for adults. (A child is not a small adult.)

The blood pH data show that the infants receiving synthetic mixtures were clearly acidotic. The low $[HCO_3^-]$ of the Neoaminosol group immediately suggests a metabolic acidosis, and the P_{CO_2} and base excess data are compatible with this interpretation. The FreAmine group, however, shows nearly normal $[HCO_3^-]$, and all of these infants have elevated P_{CO_2}'s. The P_{CO_2}'s indicate respiratory acidosis,

but a simple respiratory acidosis should be associated with a slightly elevated [HCO_3^-]. The absence of this finding in most of the infants indicates that the acidosis must also have a metabolic component. This is confirmed by the observation that all the infants receiving FreAmine have a significant negative base excess.

The infants with mixed acid–base disturbances did, in fact, have pneumonia or respiratory distress syndrome. The metabolic acidosis, which all the infants receiving synthetic mixtures experienced, was due to synthesis of aspartic acid and glutamic acid from a neutral starting material (presumably glucose). Subsequent incorporation of these acids into body protein imposed a net acid load upon the body. Addition of aspartate and/or glutamate to the synthetic mixtures was proposed as a solution of the problem.

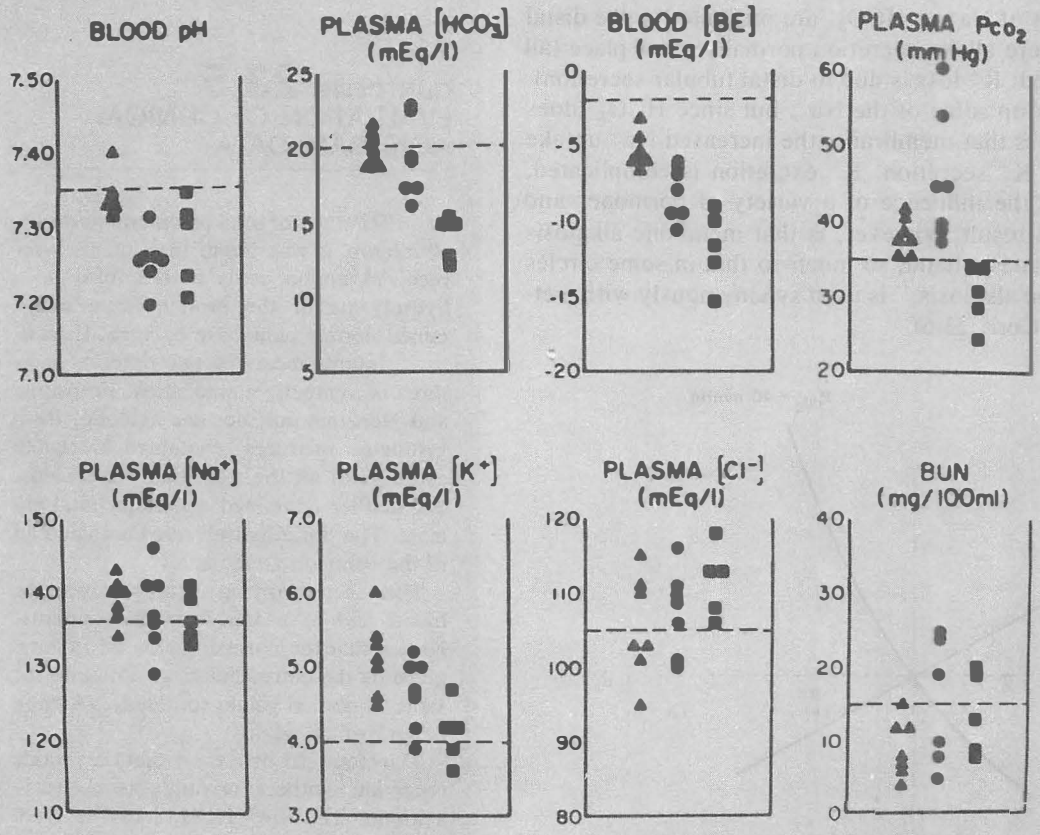


Figure C.C. 23.5

Blood acid–base data of patients receiving fibrin hydrolysate (▲) and of those receiving synthetic L-amino acid mixtures, FreAmine (●) and Neoaminosal (■).

Values are those observed at the time of the lowest blood base excess. Dashed lines represent accepted normal values for infants.

Adapted from W. C. Heird, New Engl. J. Med., 287:943, 1972.

Table C.C. 23.6

	Control	*After Lavage*	*Low KCl*	*After NaCl*
PLASMA				
pH	7.4	7.50	7.48	7.41
HCO_3^- (meq/liter)	29.3	35.3	38.1	26.1
Na^+ (meq/liter)	138	134	141	144
K^+ (meq/liter)	4.2	3.2	2.9	3.2
Cl^- (meq/liter)	101	88	85	108
URINE				
pH	6.12	7.48	5.70	7.19
HCO_3^- (meq/day)	3	51	1	17
NH_4^+ (meq/day)	22	4	36	14
Titratable acidity (meq/day)	10	0	14	1
Total acidity (meq/day)	29	−47	49	−2
Na^+ (meq/day)	2	28	1	95

SOURCE: Data from J. P. Kassirer and W. B. Schwartz, *Am. J. Med.*, 40:10, 1966.

CLIN. CORR. 23.6 METABOLIC ALKALOSIS

Prolonged gastric lavage produces a metabolic alkalosis which is a good experimental model of the metabolic alkalosis that results from repeated vomiting. The following table gives plasma and urine acid–base and electrolyte data from a healthy volunteer on a low sodium diet who, after a control period, was subjected to gastric lavage for two days. After a 5-day recovery period he was placed on a low-potassium diet and given a sodium (130 meq/day) and chloride (121 meq/day) supplement.

During the control period the data are within normal limits. After gastric lavage which selectively removed HCl (Na^+, K^+, and water lost with the gastric juice were restored), an uncomplicated metabolic alkalosis developed. Notice that the subject excreted an alkaline urine, containing a substantial amount of HCO_3^-. Na^+ excretion increased, depleting the body's Na^+ stores. Plasma P_{CO_2} was not measured, but plotting the values of pH and HCO_3^- on a pH–bicarbonate diagram (e.g., Figure 23.10) allows one to interpolate a value of about 47 mmHg. Clearly, respiratory compensation was occurring. Plasma $[K^+]$ was decreased. Plasma $[Cl^-]$ decreased, but no more than would be expected on the basis of the changes in $[Na^+]$, $[K^+]$, and $[HCO_3^-]$.

When the subject was placed on a low-potassium diet the alkalosis grew worse, and plasma $[HCO_3^-]$ rose. Additional compensatory hypoventilation evidently prevented a further rise in plasma pH. Notice, though, that the urine became acid, in spite of the increased severity of the alkalosis. Na^+ was conserved, not in

exchange for K^+, but in exchange for H^+. After several days of Na^+ and Cl^- administration, however, the subject was able to restore the depleted Cl^-, excrete the excess HCO_3^- and repair the acid–base imbalance with no other treatment.

BIBLIOGRAPHY

Gas Transport

Bunn, H. F., Gabbay, K. H., and Gallop, P. M. The glycosylation of hemoglobin: relevance to diabetes mellitus. *Science* 200:21, 1978.

Kilmartin, J. V. Interaction of haemoglobin with protons, CO_2 and 2,3-diphosphoglycerate. *Br. Med. Bull.* 32:209, 1976.

Perutz, M. F., and Lehmann, H. Molecular pathology of human haemoglobin. *Nature* 219:902, 1968.

pH Regulation

Davenport, H. W. *The ABC of acid–base chemistry*, 6th ed. Chicago: University of Chicago Press, 1974.

Gabow, P. A., Kaehny, W. D., Fennessey, P. V., et al. Diagnostic importance of an increased serum anion gap. *N. Engl. J. Med.* 303:854, 1980.

Gamble, J. L., Jr., and Bettice, J. A. Acid–base relationships in the different body compartments: the basis for a simplified diagnostic approach. *Johns Hopkins Med. J.* 140:213, 1977.

Masoro, E. J., and Siegel, P. D. *Acid–base regulation: its physiology, pathophysiology and the interpretation of blood–gas analysis*, 2nd ed. Philadelphia: W. B. Saunders, 1977.

Siggaard-Andersen, O. *The acid–base status of the blood*, 4th ed. Baltimore: Williams & Wilkins, 1974.

24

Digestion and Absorption of Basic Nutritional Constituents

ULRICH HOPFER

24.1 OVERVIEW

Historical Aspects

Secretion of digestive fluids and digestion of food were some of the earliest biochemical events to be investigated at the beginning of the era of modern science. Major concepts important for our present understanding of digestion were developed in the 19th century. Milestones were the discovery of hydrochloric acid secretion by the stomach and enzymatic hydrolysis of protein and starch by gastric juice and saliva, respectively. The discovery of gastric HCl production goes back to the American physician William Beaumont (1785–1853). In 1822 he treated a patient with a stomach wound. The patient recovered from the wound, but retained a gastric fistula (abnormal opening through the skin). Beaumont seized the opportunity to obtain and study gastric juice at different times during and after meals. Chemical analysis revealed, to the surprise of chemists and biologists, the presence of the inorganic acid HCl. This discovery established the principle of unique secretions into the gastrointestinal tract, which are elaborated by specialized glands.

Soon thereafter, the principle of enzymatic breakdown of food was recognized. Theodor Schwann, a German anatomist and physiologist (1810–1882), noticed in 1836 the ability of gastric juice to degrade albumin in the presence of dilute acid. He recognized that a new principle was involved and coined for it the word *pepsin* from the Greek *pepsis,* meaning digestion.

Today the process of secretion of digestive fluids, digestion of food, and absorption of nutrients and of electrolytes can be described in considerable detail. Progress in understanding the biochemical basis has been greatest in the areas of structure and mechanisms of digestive enzymes that are secreted as soluble proteins into the lumen of the stomach and the small intestine. In contrast, less detailed understanding is available for the more complex reactions of secretion of enzymes and of secretion as well as absorption of small solutes, encompassing both nutrients and electrolytes. Additionally, information is still incomplete on the cellular regulation of secretion and of absorption for both macromolecules and small solutes. Concepts in the less-understood areas may change in the future with increasing knowledge.

The basic nutrients fall into the classes of proteins, carbohydrates, and fats. Many different types of food can satisfy the nutritional needs of humans, even though they differ in the ratios of proteins to carbohydrates and to fats and in the ratio of digestible to nondigestible materials. Unprocessed plant products are especially rich in fibrous material that can be neither digested by human enzymes nor

Table 24.1 Contribution of Major Food Groups to Nutrient Supplies in the United States

	Total Consumption (g)	*Dairy Products, Except Butter (%)*	*Meat, Poultry, Fish (%)*	*Eggs (%)*	*Fruits, Nuts, Vegetables (%)*	*Flour, Cereal (%)*	*Sugar, Sweeteners (%)*	*Fats, Oils (%)*
Protein	100	22	42	6	12	18	0	0
Carbohydrate	381	7	0.1	0.1	19	36	37	0
Fat	155	13	35	3	4	1	0	42

easily degraded by intestinal bacteria. The fibers are mostly carbohydrates, such as cellulose (β-1,4-glucan) or pectins (mixtures of methyl esterified polygalacturic acid, polygalactose, and polyarabinose). High-fiber diets enjoy a certain popularity nowadays because of a postulated deterrent to the development of colonic cancer.

Table 24.1 describes average contributions of different food classes to the diet of North Americans. The intake of individuals may substantially deviate from the average, as food consumption depends mainly on availability and individual tastes. The ability to utilize a wide variety of food is possible because of the great adaptability and digestive reserve capacity of the gastrointestinal tract.

Knowledge of the nature of proteins and carbohydrates in the diet is important from a clinical point of view. Certain proteins and carbohydrates, although good nutrients for most humans, cannot be properly digested by some individuals and produce gastrointestinal ailments. Omission of the offending material and switching to another diet can eliminate the gastrointestinal problems for these individuals. Examples of food constituents that may be the cause of gastrointestinal disorders are gluten, one of the protein fractions of wheat, and lactose, the disaccharide in milk.

Gastrointestinal Organs and Functions

The bulk of ingested nutrients consists of large polymers that have to be broken down to monomers before they can be absorbed and made available to all cells of the body. The complete process from food intake to absorption of nutrients into the blood consists of a complicated sequence of events, of which, at the minimum, the following steps are discernible (see Figure 24.1):

1. Mechanical homogenization of food and mixing of ingested solids with fluids secreted by the glands of the gastrointestinal tract
2. Secretion of digestive enzymes that hydrolyze macromolecules to oligomers, dimers, or monomers
3. Secretion of electrolytes, acid, or base to provide an appropriate environment for optimal enzymatic digestion
4. Secretion of bile acids as detergents to solubilize lipids and facilitate their absorption
5. Hydrolysis of nutrient oligomers and dimers by intestinal surface enzymes

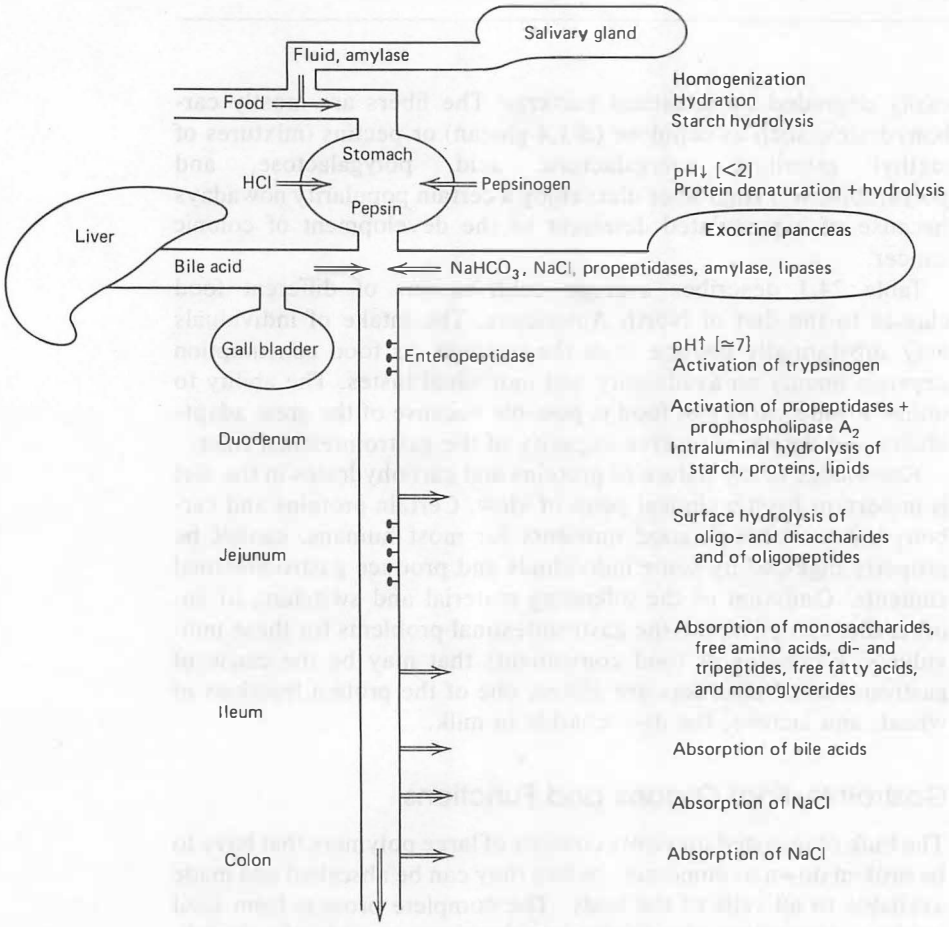


Figure 24.1
Gastrointestinal organs and their functions.

6. Transport of nutrient molecules and of electrolytes from the intestinal lumen across the epithelial cells into blood or lymph

To accomplish these diverse functions, the gastrointestinal tract contains specialized glands and surface epithelia:

Organ	*Major Function in Digestion and Absorption*
Salivary glands	Elaboration of fluid and digestive enzymes
Stomach	Elaboration of HCl and proteases
Pancreas	Elaboration of $NaHCO_3$ and enzymes for intraluminal digestion
Liver	Elaboration of bile acids
Gallbladder	Storage of bile
Small intestine	Terminal digestion of food, absorption of nutrients, and of electrolytes
Large intestine	Absorption of electrolytes

The pancreas and the small intestine are essential for digestion and absorption of all basic nutrients. Fortunately, both organs have large reserve capacities. For example, maldigestion due to pancreatic failure becomes a problem only when the pancreatic secretion rate of digestive enzymes drops below one-tenth of the normal rate. The secretion of the liver (bile) is important for efficient lipid absorption, which depends on the presence of bile acids. In contrast, gastric digestion of food is nonessential for adequate nutrition, and loss of this function can be compensated for by the pancreas and the small intestine. Yet normal gastric digestion greatly increases the smoothness and efficiency of the total digestive process. The stomach aids in the digestion through its reservoir function, its churning ability, and initiation of protein hydrolysis, which, although small, is important for stimulation of pancreatic and gallbladder output. The peptides and amino acids liberated in the stomach serve as stimuli for the coordinated release of pancreatic juice and bile into the lumen of the small intestine, thereby ensuring efficient digestion of food.

24.2 DIGESTION: GENERAL CONSIDERATIONS

Site of Digestion

Since Schwann's discovery of gastric pepsin, it has been recognized that most of the breakdown of food is catalyzed by soluble enzymes

Table 24.2 Digestive Enzymes of the Small Intestinal Surface

Enzyme (Common Name)	*Substrate*
Maltase	Maltose
Sucrase	Sucrose
Isomaltase	α-Limit dextrin
γ-Amylase	Amylose
Trehalase	Trehalose
β-Glucosidase	Glucosylceramide
Lactase	Lactose
Leucine aminopeptidase	Peptides with NH_2-terminal neutral amino acids
γ-Glutamyltransferase	Glutathione + amino acid
Enteropeptidase	Trypsinogen
Alkaline phosphatase	Orthophosphoric monoesters

and occurs within the lumen of the stomach or small intestine. However, the pancreas, not the stomach, is the major organ that synthesizes and secretes the large amounts of enzymes needed to digest the food. Secreted enzymes amount to at least 30 g protein/day in a healthy adult. The pancreatic enzymes together with bile are poured into the lumen of the second (descending) part of the duodenum, so that the bulk of the intraluminal digestion occurs distal to this site in the small intestine. However, pancreatic enzymes cannot completely digest all the nutrients to forms that can be absorbed. Even after exhaustive contact with pancreatic enzymes, a substantial portion of the carbohydrates and amino acids are present in the small intestinal lumen as dimers and oligomers, which depend for digestion on small intestinal surface or intracellular enzymes.

The importance of the small intestinal surface enzymes for digestion has been fully recognized only within the last 20 years. Its digestive functions could be examined and appreciated after methods for isolation and purification of the luminal (brush border) plasma membrane of enterocytes (intestinal epithelial cells) were developed. This membrane contains on its extracellular side many di- and oligosaccharidases, amino- and dipeptidases, as well as esterases (Table 24.2). Many of these enzymes protrude up to 100 Å into the intestinal lumen, attached to the plasma membrane by an anchoring polypeptide that itself has no role in the hydrolytic activity. The substrates for these enzymes are the nutrient oligomers and dimers that result from the pancreatic digestion of food. The surface enzymes are glycoproteins that are relatively stable toward digestion by pancreatic proteases or toward detergents.

A third site of digestion is the cytoplasm of enterocytes. Intracellular digestion is of some importance for the hydrolysis of di- and tripeptides, which can be absorbed across the luminal plasma membrane.

Exocrine Secretion of Enzymes

The salivary glands, the gastric mucosa, and the pancreas contain specialized cells for the synthesis, packaging, and release of enzymes into the lumen of the gastrointestinal tract. (See Figure 24.2.) This secretion is termed "exocrine" because of its direction toward the lumen.

Proteins destined for secretion are synthesized on the polysomes of the rough endoplasmic reticulum, which is particularly abundant in exocrine cells. The NH_2-terminal amino acid sequence from nascent secretory proteins constitutes a signal, which causes release of the NH_2-terminus into the cisternal space of the endoplasmic reticulum. With further synthesis, the entire polypeptide is fed through

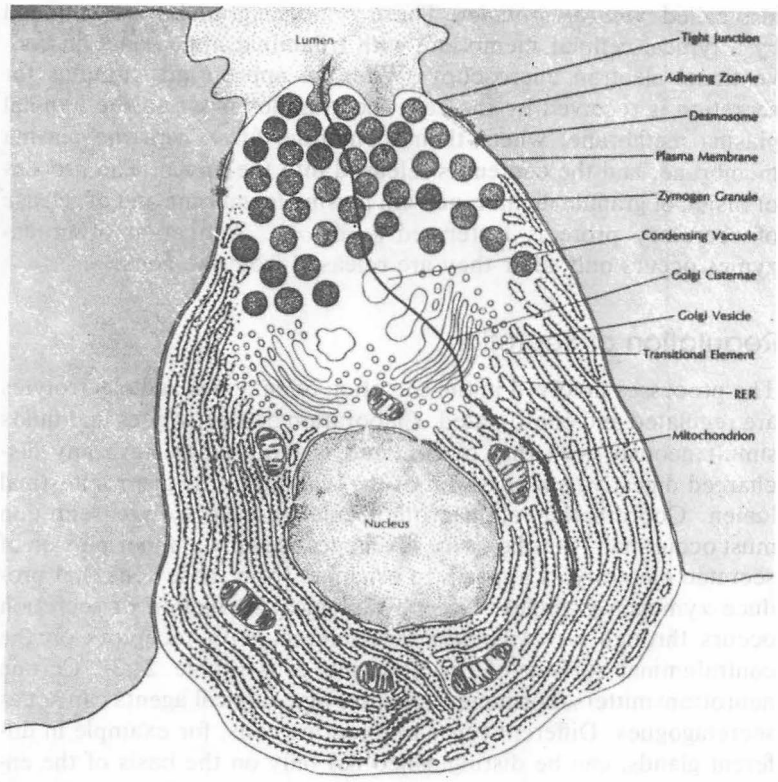


Figure 24.2
Exocrine secretion of digestive enzymes.
Reproduced with permission from J. D. Jamieson, Membrane and secretion, in Cell Membranes: biochemistry, cell biology and pathology, G. Weissmann and R. Claiborne, eds., HP Publishing Co., Inc., New York, 1975. Figure by B. Tagawa.

the endoplasmic membrane so that the final product is sequestered within the endoplasmic cisterna. The amino acids forming the signal sequence may be clipped off during further processing. Secretory proteins are then transported from the endoplasmic reticulum to the Golgi complex in small membrane-bound vesicles. In the endoplasmic reticulum and also the Golgi apparatus, glycosylation may occur. Subsequent to processing, the secretory proteins are packaged into larger vesicles ~1 μm in diameter, which serve as storage form until the stimulus for secretion is received. Proteases and phospholipase A are produced and stored as inactive precursors, also termed proenzymes or zymogens. Therefore the storage vesicles are

also called *zymogen granules*. These zymogen granules are bounded by a typical cellular membrane with trilaminar appearance in conventional electron microscopy. When an appropriate stimulus for secretion is received by the cell, the granules move to the luminal plasma membrane, where their membrane fuses with the plasma membrane, and the content is released into the lumen. The process of fusion of granule membrane with plasma membrane and of release of secretory proteins is termed *exocytosis*. Activation of proenzymes occurs only after they are released from the cells.

Regulation of Secretion

The processes involved in the secretion of enzymes and electrolytes are regulated and coordinated. Elaboration of electrolytes and fluids simultaneously with that of enzymes is required to flush any discharged digestive enzymes out of the gland into the gastrointestinal lumen. Coordination of macromolecule and electrolyte secretion must occur mainly at the tissue level because only a minor portion of secreted electrolytes appears to originate from those cells that produce zymogen granules. The physiological regulation of secretion occurs through *secretagogues* that interact with receptors on the contraluminal surface of the exocrine cells (Table 24.3). Certain neurotransmitters, hormones, and pharmacological agents can act as secretagogues. Different types of exocrine cells, for example in different glands, can be distinguished not only on the basis of the enzyme secreted but also with respect to the types of receptors and hence the agents that conserve secretagogues. Ca^{2+} as well as cAMP have been implicated as intracellular messengers of secretagogues, although the exact mechanism of stimulation–secretion is not yet completely understood. Some secretagogues are stimulators of adenylate cyclase and raise intracellular levels of cAMP, which can be correlated with secretory states; other secretagogues increase intracellular Ca^{2+} levels without changing cAMP. (see Figure 24.3.)

Acetylcholine elicits salivary, gastric, and pancreatic enzyme and electrolyte secretion. (See Figure 24.4.) It appears to be the major neurotransmitter for stimulating secretion, with input from the central nervous system in salivary and gastric glands, or via local reflexes in gastric glands and the pancreas. The acetylcholine receptor of exocrine cells is of the *muscarinic* type, that is, it can be blocked by atropine (Figure 24.5). Most people have experienced the effect of atropine because it is used by dentists to "dry up" the mouth for dental work.

A second class of secretagogues consists of certain biogenic amines. For example, histamine is a potent stimulator of HCl secretion. (See Figure 24.6.) It interacts with a gastric-specific histamine

Table 24.3 Physiological Secretagogues

Organ	*Secretion*	*Secretagogue*
Salivary gland	NaCl, amylase	Acetylcholine (catecholamines?
Stomach	HCl, pepsinogen	Acetylcholine, histamine, gastrin
Pancreas	NaCl, enzymes	Acetylcholine, cholecystokinin (secretin)
	$NaHCO_3$,NaCl	Secretin

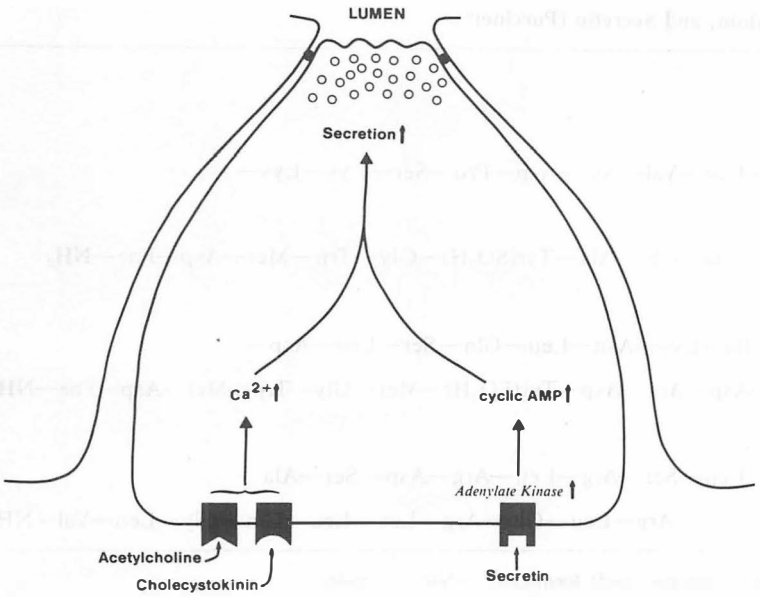


Figure 24.3
Cellular regulation of exocrine secretion in the pancreas.
Reproduced with permission from J. D. Gardner, Annu. Rev. Physiol. 41:63, 1979. Copyright 1979 by Annual Reviews Inc.

receptor (=H_2 receptor) on the contraluminal plasma membrane of oxyntic cells. The cellular origin of histamine involved in the regulation of HCl secretion is not exactly known. Histamine as well as analogs, which act as antagonists at the H_2 receptor, are used medically to increase or decrease HCl output. 5-Hydroxytryptamine (serotonin) is another biogenic amine that is present in relatively high amounts in the gastrointestinal tract (Figure 24.7). It probably is involved in stimulation of NaCl secretion by the small intestinal mucosa.

A third class of secretagogues consists of peptide hormones. The gastrointestinal tract is rich in specialized epithelial cells, containing a large number of different biologically active amines or peptides. The peptides are localized in granules, usually close to the contraluminal pole of these cells and probably are released into the interstitial space. Hence these epithelial cells are classified as endocrine cells. Of particular importance are the peptides gastrin, cholecystokinin (pancreozymin), and secretin (Table 24.4).

Gastrin occurs predominantly as either a large peptide of 34 amino acids (G-34) or a smaller one of 17 residues (G-17) from the COOH-terminus of G-34. The functional portion of gastrin resides mainly in

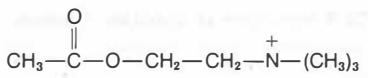


Figure 24.4
Acetylcholine.

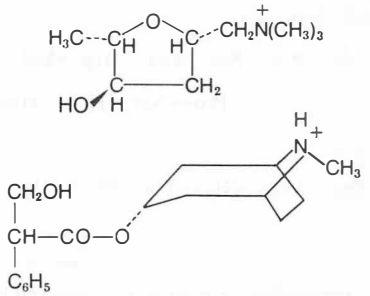


Figure 24.5
L(+)-Muscarine (top) and atropine (bottom).

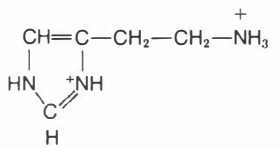


Figure 24.6
Histamine.

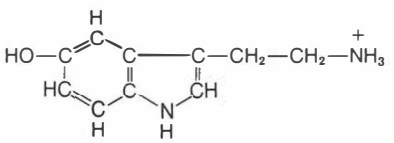


Figure 24.7
5-OH-Tryptamine (serotonin).

Table 24.4 Structure of Gastrin (Human), Cholecystokinin, and Secretin (Porcine)[a]

GASTRIN

G-34
|
Glp—Leu—Gly—Pro—Gln—Gly—His—Pro—Ser—Leu—Val—Ala—Asp—Pro—Ser—Lys—Lys—

G-17
|
Gln—Gly—Pro—Trp—Leu—Glu—Glu—Glu—Glu—Glu—Ala—Tyr(SO_3H)—Gly—Trp—Met—Asp—Phe—NH_2

CHOLECYSTOKININ

Lys—Ala—Pro—Ser—Gly—Arg—Val—Ser—Met—Ile—Lys—Asn—Leu—Gln—Ser—Leu—Asp—
Pro—Ser—His—Arg—Ile—Ser—Asp—Arg—Asp—Tyr(SO_3H)—Met—Gly—Trp—Met—Asp—Phe—NH_2

SECRETIN

His—Ser—Asp—Gly—Thr—Phe—Thr—Ser—Glu—Leu—Ser—Arg—Leu—Arg—Asp—Ser—Ala—
Arg—Leu—Gln—Arg—Leu—Leu—Gln—Gly—Leu—Val—NH_2

[a] Glp = pyrrolidino carboxylic acid, derived from Glu through internal amide formation; —NH_2 = amide.

the last 5 amino acids of the COOH-terminus. Thus pentagastrin, an artificial pentapeptide containing only the last five amino acids, can be used specifically to stimulate gastric HCl and pepsin secretion. However, the additional amino acids in the natural gastrins increase the potency of the hormone. Gastrin as well as cholecystokinin have an interesting chemical feature, a sulfated tyrosine, which considerably enhances the potency of both hormones.

Cholecystokinin and pancreozymin denote the same peptide. The different names stem from the times when only the functions were known before the peptides had been purified. As implied by the names, the peptide stimulates gallbladder contraction and secretion of pancreatic enzymes. The peptide is secreted by epithelial endocrine cells of the small intestine, particularly in the duodenum. Cholecystokinin secretion is stimulated by luminal amino acids and peptides, usually derived from gastric proteolysis, by fatty acids and an acid pH. Cholecystokinin and gastrin are thought to be related in an evolutionary sense, as both share an identical amino acid sequence at the COOH-terminus.

Secretin is a polypeptide of 27 amino acids. This peptide is secreted by yet other endocrine cells of the small intestine. Its secretion is stimulated particularly by luminal pH <5. The major biological activity of secretin is stimulation of pancreatic juice rich in $NaHCO_3$. Pancreatic $NaHCO_3$ is essential for the neutralization of gastric HCl when it reaches the duodenum. Secretin also enhances

$$HCl + NaHCO_3 \longrightarrow NaCl + H_2O + CO_2$$

pancreatic enzyme release, acting synergistically with cholecystokinin.

24.3 EPITHELIAL TRANSPORT

General Considerations

Solute movement across an epithelial cell layer is determined by the properties of epithelial cells, particularly their plasma membranes, as well as by the intercellular tight junctional complexes. (See Figure 24.8.) The tight junctions extend in a beltlike manner around the perimeter of each epithelial cell and connect neighboring cells. Therefore, the tight junctions constitute part of the barrier between the two extracellular spaces on either side of the epithelium, that is, the lumen of the gastrointestinal tract and the intercellular (interstitial) space on the blood or nutrient side. The tight junction also marks the boundary between the luminal and the contraluminal region of the plasma membrane of epithelial cells.

Two potential parallel pathways for solute transport across epithelial cell layers can be distinguished: through the cells (transcellular) and through the tight junctions between cells (paracellular) (Figure 24.8). The transcellular route in turn consists mainly of two barriers in series, which are formed by the luminal and by the contraluminal

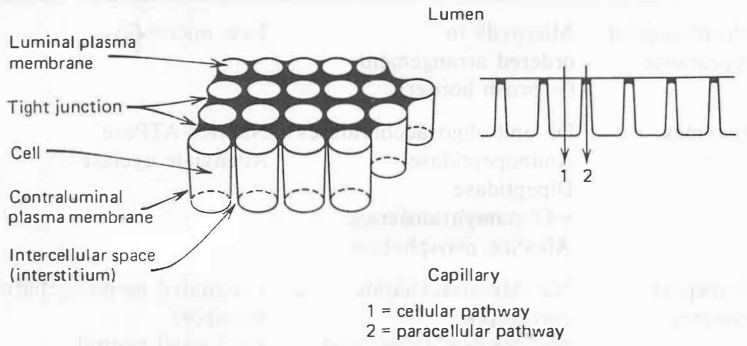


Figure 24.8
Pathways for transport across epithelia.

plasma membrane. Because of this combination of different barriers in parallel (cellular and paracellular pathways) and in series (luminal and contraluminal plasma membranes), biochemical and biophysical information on all three barriers as well as their mutual influence is required for understanding the overall transport properties of the epithelium.

One of the main functions of epithelial cells in the gastrointestinal tract is active transport of nutrients, electrolytes, and vitamins from one side of the epithelium to the other. The cellular basis for this vectorial solute movement must lie in the different properties of the luminal and contraluminal regions of the plasma membrane. The small intestinal cells provide a prominent example of the differentiation and specialization of the two types of membrane. They differ in morphological appearance, enzymatic composition, chemical composition, and transport functions (Table 24.5). The luminal membrane in contact with the nutrients in the chyme (the semifluid mass of partially digested food) is specialized for terminal digestion of nutrients through its digestive enzymes and for nutrient absorption through transport systems that accomplish concentrative uptake. Such transport systems are well known for monosaccharides, amino acids, peptides, or electrolytes. In contrast, the contraluminal plasma membrane, which is in contact with the intercellular fluid, capillaries, and lymph, has properties similar to the plasma membrane of most cells. It possesses receptors for hormonal or neuronal

Table 24.5 Characteristic Differences Between Luminal and Contraluminal Plasma Membrane of Small Intestinal Epithelial Cells

	Luminal	*Contraluminal*
Morphological appearance	Microvilli in ordered arrangement (= brush border)	Few microvilli
Enzymes	Di- and oligosaccharidases Aminopeptidase Dipeptidase γ-Glutamyltransferase Alkaline phosphatase	Na^+,K^+-ATPase Adenylate cyclase
Transport systems	Na^+-Monosaccharide cotransport Na^+-Neutral amino acid cotransport Na^+-Bile acid cotransport	Facilitated monosaccharide transport Facilitated neutral amino acid transport

regulation of cellular functions, a Na^+,K^+-ATPase for removal of Na^+ from the cell and transport systems for the entry of nutrients for its own consumption. In addition, the contraluminal plasma membrane contains the transport systems necessary for exit of the nutrients absorbed from the lumen so that the digested food can become available to all cells of the body. Some of the transport systems in the contraluminal plasma membrane may fulfill both the function of catalyzing exit when the intracellular nutrient concentration is high after a meal and that of mediating their entry when the blood levels are higher than those within the cell.

NaCl Absorption

The transport of Na^+ plays a crucial role not only for epithelial NaCl absorption or secretion, but also in the energization of nutrient uptake. The Na^+,K^+-ATPase provides the dominant mechanism for transduction of chemical energy in the form of ATP into osmotic energy of a concentration (chemical) or a combined concentration and electrical (electrochemical) ion gradient across the plasma membrane. In epithelial cells this enzyme is located exclusively in the contraluminal plasma membrane (Figure 24.9). The best estimate for the stoichiometry of the Na^+,K^+-ATPase reaction appears to be 1 mol ATP coupled to the outward pumping of 3 mol Na^+ and the simultaneous inward pumping of 2 mol K^+. Because three cations leave the cell, while only two enter it during each cycle of ATP hydrolysis, the cytoplasm becomes electrically negative with respect to the extracellular fluid. In small intestinal epithelial cells, intracellular Na^+ concentrations are about 30 mM, that is, fivefold lower than in plasma, and the cytoplasm is about -60 mV relative to the extracellular solution.

Transepithelial NaCl movements are produced by the combined actions of the Na^+,K^+-ATPase and additional transport systems in the plasma membrane, which allow the entry of Na^+ or Cl^- into the cell. If NaCl can enter the cell passively at the luminal pole, NaCl absorption results from the combined action of NaCl entry and of Na^+,K^+-ATPase activity; if NaCl can enter the cell passively through the contraluminal plasma membrane, NaCl is secreted into the lumen.

At least two different types of transport systems for "passive" Na^+ movements across the luminal plasma membrane can be distinguished on the basis of inhibitors, hormonal regulation, and electrical fluxes associated with Na^+ transport. The epithelial cells of the lower portion of the large intestine possess a transport system that allows the uncoupled entry of Na^+ into the cell down its electrochemical gradient (Figure 24.10). This Na^+ flux is associated with

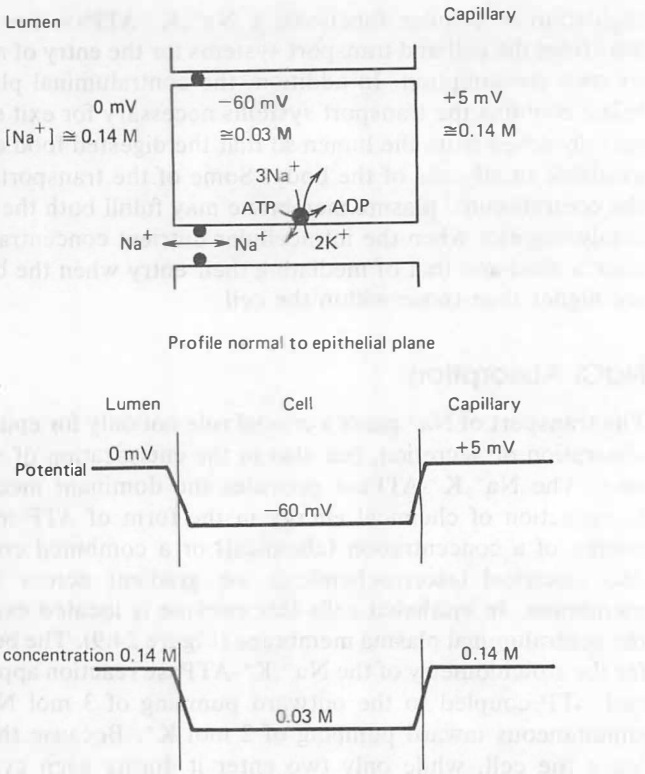


Figure 24.9
Na^+ concentrations and electrical potentials in enterocytes.

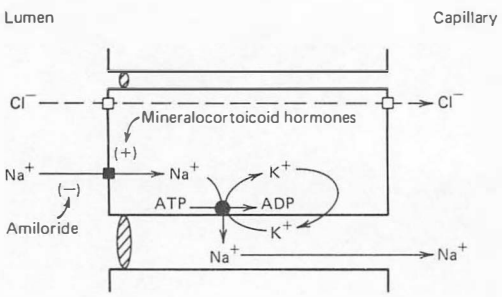


Figure 24.10
Model for epithelial NaCl transport in the lower intestine.

an electrical current, that is, electrogenic, and can be inhibited by the drug amiloride at micromolar concentrations (Figure 24.11). The presence of this transport system, and hence NaCl absorption, is regulated by mineralocorticoid hormones. In contrast, epithelial cells of the upper portion of the intestine possess a transport system in the brush border membrane, which catalyzes an electrically silent Na^+/H^+ exchange. (See Figure 24.12.) The exchange is not affected by low concentrations of amiloride and not regulated by mineralocorticoids. In both cases, the location and the function of the Na^+,K^+-ATPase is identical, namely, to extrude Na^+ at the contraluminal pole. The necessity for two types of NaCl absorption may arise from the different functions of upper and lower intestine, which require different regulation. The upper intestine reabsorbs the bulk of NaCl from the diet and the secretions of the exocrine glands after each meal, and the lower intestine participates in the fine regula-

tion of NaCl retention, depending on the overall electrolyte balance of the body.

NaCl and $NaHCO_3$ Secretion

The gastrointestinal tract contains many different epithelia that have the ability to secrete electrolytes. The major ones are Na^+ and Cl^-. Water follows passively because of the osmotic forces exerted by any secreted solute. Thus, NaCl secretion secondarily results in fluid secretion. The fluid may be either hypertonic or isotonic, depending on the contact time of the secreted fluid with the epithelium and the tissue permeability to water. The longer the contact and the greater the water permeability, the better osmotic equilibrium, that is, isotonicity, is achieved. (See Figure 24.13.)

The cellular mechanisms for NaCl secretion are not completely established, but appear to involve the Na^+,K^+-ATPase located in the contraluminal plasma membrane of epithelial cells (Figure 24.14). The enzyme is implicated because cardiac glycosides, inhibitors of this enzyme, abolish salt secretion. However, the involvement of Na^+,K^+-ATPase does not provide a straightforward explanation for a NaCl movement from the capillary side to the lumen because the enzyme extrudes Na^+ from the cell toward the capillary side. Thus the active step of Na^+ transport across one of the plasma membranes has a direction opposite to that of overall transepithelial NaCl movements. The current model of NaCl secretion, which is supported mainly by data from the salt-secreting glands of birds and sharks, has solved the paradox by an electrical coupling of Cl^- secretion across the luminal plasma membrane and Na^+ movements via the paracellular route illustrated in Figure 24.14.

In the pancreas a fluid rich in Na^+ and Cl^- is secreted by acinar cells. This fluid provides the vehicle for the movement of digestive enzymes from the acini, where they are released, to the lumen of the duodenum. As the fluid passes through the ducts, it is modified by an exchange of Cl^- for HCO_3^-. The HCO_3^- concentration in the final pancreatic juice can reach concentrations of up to 120 mM. (See Figure 24.15.)

The permeability of the tight junction to water, Na^+, or other ions modifies active transepithelial solute movements. For example, a high permeability to Na^+ allows this cation to equilibrate between extracellular solutions in contact with contraluminal and luminal plasma membranes. Thus the electrochemical Na^+ gradient, established by the Na^+,K^+-ATPase across the contraluminal plasma membrane, can be transmitted to the luminal plasma membrane, where it is available to energize the transport of other solutes. Different regions of the gastrointestinal tract differ not only with respect

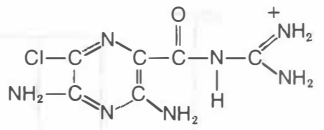


Figure 24.11
Amiloride.

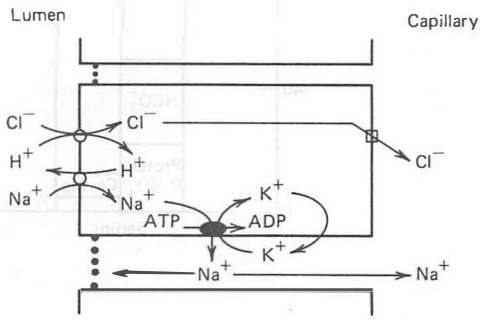


Figure 24.12
Model for epithelial NaCl transport in the upper intestine.

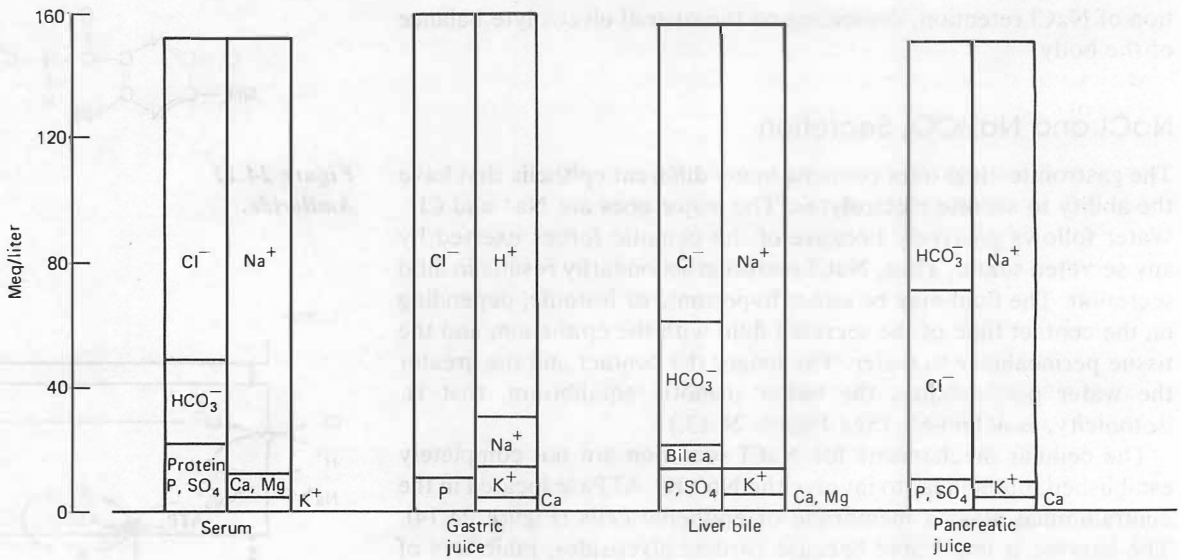


Figure 24.13
Ionic composition of secretions of the gastrointestinal tract.
Serum is included to facilitate comparison between fluids. Note the high H^+ concentration in gastric juice ($pH < 1$) and the high HCO_3^- concentration in pancreatic juice. P, organic and inorganic phosphate, SO_4, inorganic and organic sulfate, Ca, calcium, Mg, magnesium, bile a., bile acids.
Adapted from *Biological Handbooks, Blood and Other Body Fluids*, Federation of American Societies for Experimental Biology, 1961.

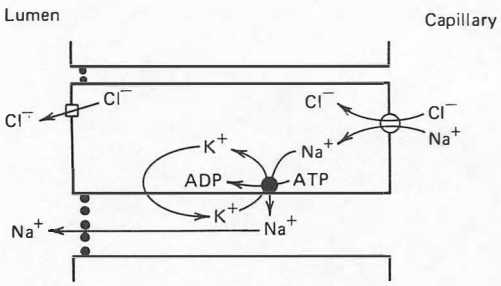


Figure 24.14
Model for epithelial NaCl secretion.

to the transport systems that determine the passive membrane permeability (see above for amiloride-sensitive and amiloride-insensitive Na^+ entry), but also with respect to the permeability characteristics of the tight junction.

Energization of Nutrient Transport

Many solutes are absorbed across the intestinal epithelium against a concentration gradient. The energy for this "active" transport is directly derived from the Na^+ concentration gradient or the electrical potential across the luminal plasma membrane, rather than from the chemical energy of a covalent bond change, such as ATP hydrolysis. Glucose transport provides an example of uphill solute transport that is driven directly by the electrochemical Na^+ gradient and only indirectly by ATP (Figure 24.16).

In vivo, glucose is absorbed from the lumen into the blood against a concentration gradient. This vectorial transport is the combined result of several separate membrane events: (1) the asymmetric insertion of different transport systems for glucose into the luminal and the contraluminal plasma membrane, (2) the coupling of Na^+ and glucose transport across the luminal membrane, and (3) the ATP-dependent Na^+ transport out of the cell at the contraluminal pole. (See Figure 24.17.)

The luminal plasma membrane contains a transport system that facilitates a tightly coupled movement of Na^+ and D-glucose (or structurally similar sugars). The transport system mediates glucose and Na^+ transport equally well in both directions. However, because of the higher Na^+ concentration in the lumen and the negative potential within the cell, the observed reaction is from lumen to cell, even if the cellular glucose concentration is higher than the luminal one. In other words, downhill Na^+ movement normally supports concentrative glucose transport. Concentration ratios of up to 20-fold between intracellular and extracellular glucose have been observed in vitro under conditions of blocked efflux of cellular glucose.

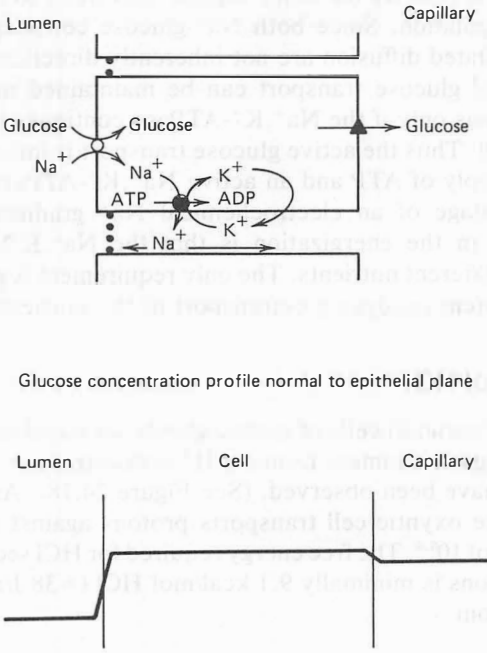


Figure 24.16
Model for epithelial glucose transport.

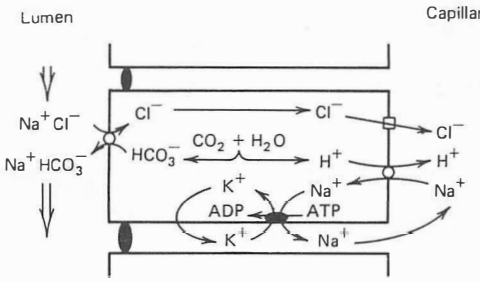


Figure 24.15
Model for epithelial HCO_3^- secretion.

$$3Na^+_{lumen} + 3Glc_{lumen} \rightleftharpoons 3Na^+_{cell} + 3Glc_{cell}$$

$$3Na^+_{cell} + 2K^+_{interstitium} + ATP_{cell} \dashrightarrow 3Na^+_{interstitium} + 2K^+_{cell} + ADP_{cell} + P_{cell}$$

$$2K^+_{cell} \rightleftharpoons 2K^+_{interstitium}$$

$$3Na^+_{interstitium} \rightleftharpoons 3Na^+_{lumen}$$

$$3Glc_{cell} \rightleftharpoons 3Glc_{interstitium}$$

$$\text{Sum:} \quad 3Glc_{lumen} + ATP_{cell} \longrightarrow 3Glc_{interstitium} + ADP_{cell} + P_{cell}$$

Figure 24.17
Transepithelial glucose transport as translocation reactions across the plasma membranes and the tight junction.

The contraluminal plasma membrane contains another type of transport system for glucose, which allows glucose to exit. This transport system facilitates equilibration of glucose across the membrane, whereby the direction of the net flux is determined by the glucose concentration gradient. The two glucose transport systems in the luminal and contraluminal plasma membrane share glucose as substrate, but otherwise differ considerably in terms of Na^+ as cosubstrate, specificity for other sugars, sensitivity to inhibitors, or biological regulation. Since both Na^+-glucose cotransport and the simple, facilitated diffusion are not inherently directional, "active" transepithelial glucose transport can be maintained under steady-state conditions only if the Na^+,K^+-ATPase continues to move Na^+ out of the cell. Thus the active glucose transport is indirectly dependent on a supply of ATP and an active Na^+,K^+-ATPase.

The advantage of an electrochemical Na^+ gradient serving as intermediate in the energization is that the Na^+,K^+-ATPase can drive many different nutrients. The only requirement is presence of a transport system catalyzing cotransport of the nutrient with Na^+.

Secretion of HCl

The oxyntic (parietal) cells of gastric glands are capable of secreting HCl into the gastric lumen. Luminal H^+ concentrations of up to 0.15 M (pH 0.8) have been observed. (See Figure 24.18.) As the plasma pH = 7.4, the oxyntic cell transports protons against a concentration gradient of $10^{6.6}$. The free energy required for HCl secretion under these conditions is minimally 9.1 kcal/mol HCl (=38 J/mol HCl), as calculated from

$$\Delta G' = RT\ 2.3 \log 10^{6.6} \qquad (RT = 0.6 \text{ kcal/mol at } 37°C)$$

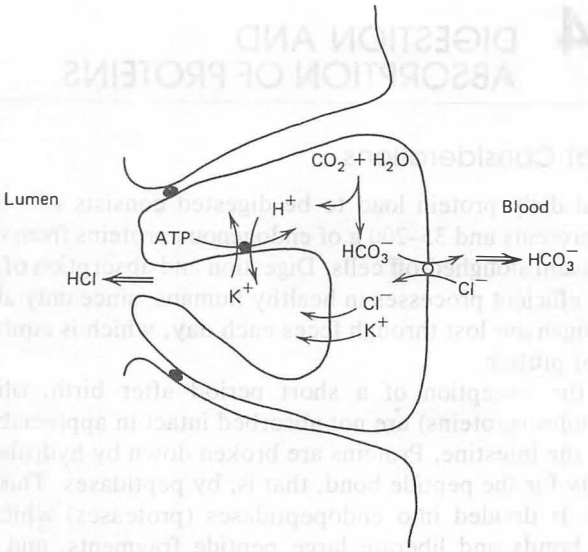


Figure 24.18
Model for secretion of hydrochloric acid.

A K^+-activated ATPase is intimately involved in the mechanism of active HCl secretion. This enzyme is unique to the oxyntic cell and is found only in the luminal region of the plasma membrane. It couples the hydrolysis of ATP to an electrically silent obligatory exchange of K^+ for H^+, secreting H^+ and taking K^+ into the cell. The stoichiometry appears to be 1 mol of transported H^+ and K^+ for each mol of ATP:

$$ATP_{cell} + H^+_{cell} + K^+_{lumen} \rightleftharpoons ADP_{cell} + P_{cell} + H^+_{lumen} + K^+_{cell}$$

In the steady state, HCl can be elaborated by this mechanism only if the luminal membrane is permeable to K^+ and Cl^- and the contraluminal plasma membrane catalyzes an exchange of Cl^- for HCO_3^-. The exchange of Cl^- for HCO_3^- is essential to refurnish the cell with Cl^- and to prevent accumulation of base within the cell. Thus, under steady-state conditions, secretion of HCl into the gastric lumen is coupled to that of HCO_3^- into the plasma.

24.4 DIGESTION AND ABSORPTION OF PROTEINS

General Considerations

The total daily protein load to be digested consists of ~70–100 g dietary proteins and 35–200 g of endogenous proteins from digestive enzymes and sloughed-off cells. Digestion and absorption of proteins are very efficient processes in healthy humans, since only about 1–2 g of nitrogen are lost through feces each day, which is equivalent to 6–12 g of protein.

With the exception of a short period after birth, oligo- and polypeptides (proteins) are not absorbed intact in appreciable quantities by the intestine. Proteins are broken down by hydrolases with specificity for the peptide bond, that is, by peptidases. This class of enzymes is divided into endopeptidases (proteases) which attack internal bonds and liberate large peptide fragments, and exopeptidases which cleave off one amino acid at a time from either the COOH- or the NH_2-terminus. Thus exopeptidases are further subdivided into carboxy- and aminopeptidases. Endopeptidases are important for an initial breakdown of long polypeptides into smaller products, which can then be attacked by the exopeptidases. In effect, the action of endopeptidases increases the substrate concentration for exopeptidases.

The process of protein digestion can be divided into a gastric, a pancreatic, and an intestinal phase, depending on the source of peptidases. (See Figure 24.19.)

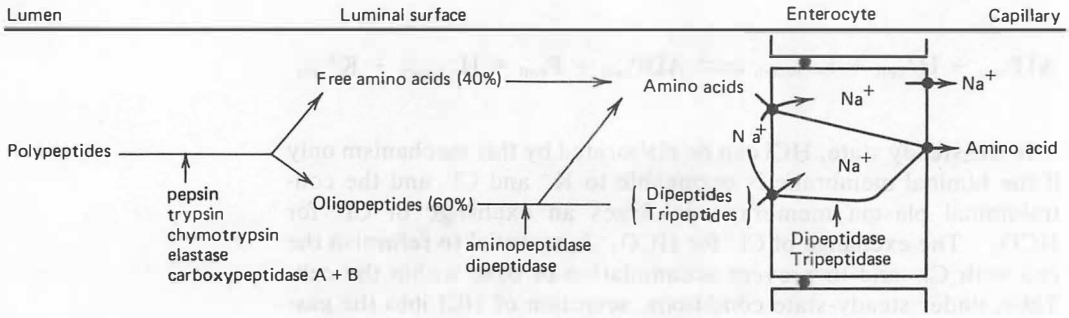


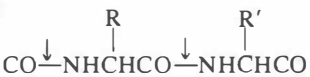
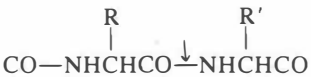
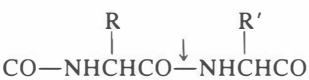
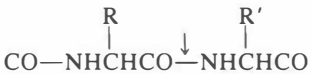
Figure 24.19
Digestion and absorption of proteins.

Gastric Digestion of Proteins

Gastric juice is characterized by the presence of HCl and therefore a low pH <2 as well as the presence of proteases of the pepsin family. The acid serves to kill off microorganisms and also to denature proteins. Denaturation makes proteins more susceptible to hydrolysis by proteases. Pepsins are unique in that they are acid-stable; in fact, they are active at acid but not at neutral pH. The catalytic mechanism that is effective for peptide hydrolysis at the acid pH depends on two carboxylic groups at the active site of the enzymes. Pepsin A, the major gastric protease, prefers peptide bonds formed by the amino group of aromatic amino acids (Phe, Tyr). (See Table 24.6.)

Active pepsin is generated from the proenzyme pepsinogen by the removal of 44 amino acids from the NH_2-terminus (pig enzyme). Cleavage of the peptide bond between residues 44 and 45 of pepsinogen can occur as either an intramolecular reaction (autoactivation) below pH 5 or by active pepsin (autocatalysis). The liberated peptide

Table 24.6 Gastric and Pancreatic Peptidases

Enzyme	*Proenzyme*	*Activator*	*Reaction Catalyzed*	
Carboxyl Proteases				
Pepsin A	Pepsinogen A	Autoactivation, pepsin		R = tyr, phe, leu
Serine Proteases				
Trypsin	Trypsinogen	Enteropeptidase, trypsin		R = arg, lys
Chymotrypsin	Chymotrypsinogen	Trypsin		R = tyr, trp, phe, met, leu
Elastase	Proelastase	Trypsin		R = ala, gly, ser
Zn-Peptidases				
Carboxypeptidase A	Procarboxypeptidase A	Trypsin		R = val, leu, ile, ala
Carboxypeptidase B	Procarboxypeptidase B	Trypsin		R = arg, lys

from the NH_2-terminus remains bound to pepsin and acts as "pepsin inhibitor" above pH 2. This inhibition is released either by a drop of the pH below 2 or further degradation of the peptide by pepsin. Thus, once favorable conditions are reached, pepsinogen is converted to pepsin by autoactivation and subsequent autocatalysis at an exponential rate.

The major products of pepsin action are large peptide fragments and a few free amino acids. The importance of gastric protein digestion does not lie so much in its contribution to the breakdown of ingested macromolecules, but rather in the generation of peptides and amino acids that act as stimulants for cholecystokinin release in the duodenum. The gastric peptides therefore are instrumental in the initiation of the pancreatic phase of protein digestion.

Pancreatic Digestion of Proteins

The pancreatic juice is rich in proenzymes of endopeptidases and carboxypeptidases. (See Figure 24.20.) These proenzymes are activated only after they reach the lumen of the small intestine. The key to activation is enteropeptidase (old name: enterokinase), a protease produced by duodenal epithelial cells. Enteropeptidase activates pancreatic trypsinogen to trypsin by scission of a hexapeptide from the NH_2-terminus. Trypsin in turn autocatalytically activates more trypsinogen to trypsin and also acts on the other proenzymes, thus

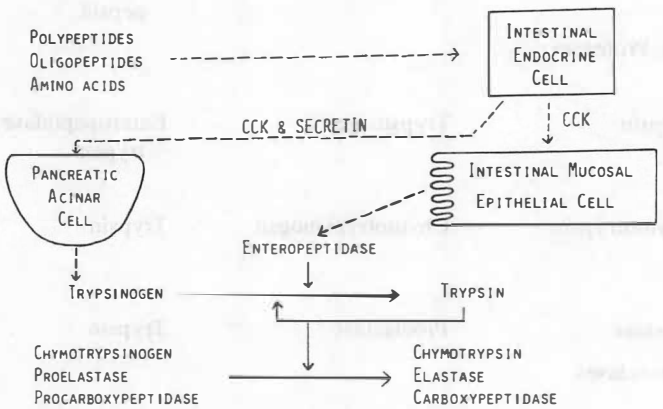


Figure 24.20
Secretion and activation of pancreatic enzymes.
Abbreviation: CCK = cholecystokinin.
Reproduced with permission from H. J. Freeman and Y. S. Kim, Annu. Rev. Med. 29:102, 1978. Copyright 1978 by Annual Reviews Inc.

liberating the endopeptidases chymotrypsin and elastase and the carboxypeptidases A and B. Since trypsin plays a pivotal role among pancreatic enzymes in the activation process, pancreatic juice normally contains a small molecular weight peptide that acts as a trypsin inhibitor and neutralizes any trypsin formed prematurely within the pancreatic cells or pancreatic ducts.

Trypsin, chymotrypsin, and elastase are endopeptidases with different substrate specificity as shown in Table 24.6. They are all active only at neutral pH and depend on pancreatic $NaHCO_3$ for neutralization of gastric HCl. The mechanism of catalysis of all three enzymes involves an essential "serine" residue. Thus reagents that interact with serine and modify it, inactivate the enzymes. A prominent example of such a reagent is diisopropylphosphofluoridate (Figure 24.21).

The polypeptides generated from ingested proteins by the action of gastric and pancreatic endopeptidases are degraded further within the small intestinal lumen by carboxypeptidase A and B. The pancreatic carboxypeptidases are metalloenzymes that require Zn^{2+} for activity and thus possess a different type of catalytic mechanism than the "carboxy" or "serine" peptidases.

The combined action of pancreatic peptidases results in the formation of free amino acids and small peptides of 2–8 residues. Peptides account for about 60% of the amino nitrogen at this point.

Small Intestinal Digestion of Proteins

Since pancreatic juice does not contain appreciable aminopeptidase activity, final digestion of di- and oligopeptides depends on small intestinal enzymes. The luminal surface of intestinal epithelial cells is particularly rich in aminopeptidase activity, but also contains di-

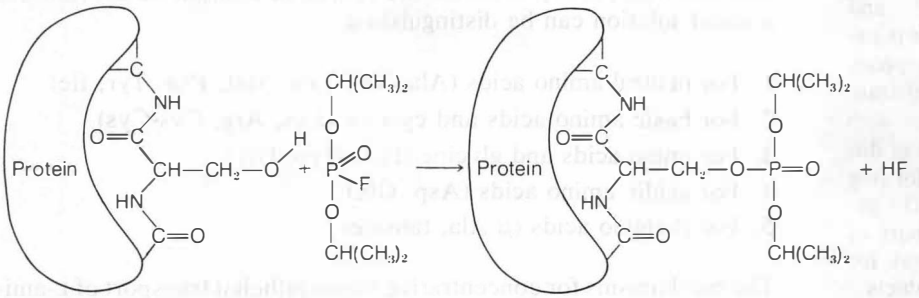


Figure 24.21
Inactivation of "serine"-proteases by diisopropylphosphofluoridate.

peptidases. The end products of the cell surface digestion are free amino acids and di- and tripeptides. Amino acids and small peptides are then absorbed by the epithelial cells via specific amino acid or peptide transport systems. The di- and tripeptides are generally hydrolyzed within the cytoplasmic compartment before they leave the cell. The cytoplasmic dipeptidases explain why practically only free amino acids are found in the portal blood after a meal. The virtual absence of peptides had previously been taken as evidence that luminal protein digestion had to proceed all the way to free amino acids before absorption could occur. However, it is now established that a large portion of dietary amino nitrogen is absorbed in the form of small peptides with subsequent intracellular hydrolysis. Exception to this general rule are di- and tripeptides containing proline and hydroxyproline or unusual amino acids, such as β-alanine in carnosine (β-alanylhistidine) or anserine [β-alanyl(1-methyl)histidine] after ingestion of chicken meat. These peptides are not good substrates for the intestinal cytoplasmic dipeptidases and therefore are available for transport out of the cell into the portal blood.

CLIN. CORR. 24.1 NEUTRAL AMINO ACIDURIA (HARTNUP DISEASE)

Transport functions, like enzymatic functions, are subject to modification by mutations. An example of a genetic lesion in epithelial amino acid transport is Hartnup disease, named after the family in which the disease entity resulting from the defect was first recognized. The disease is characterized by the inability of renal and intestinal epithelial cells to absorb neutral amino acids from the lumen. In the kidney, in which plasma amino acids reach the lumen of the proximal tubule through the ultrafiltrate, the inability to reabsorb amino acids manifests itself as excretion of amino acids in the urine (amino aciduria). The intestinal defect results in malabsorption of free amino acids from the diet. Therefore the clinical symptoms of patients with this disease are mainly those due to essential amino acid and nicotinamide deficiencies. The latter is explained by a deficiency of tryptophan, which serves as precursor for nicotinamide. Investigations of patients with Hartnup disease revealed existence of different intestinal transport systems for free amino acids and di- or tripeptides. The genetic lesion does not affect transport of peptides which remain as a pathway for absorption of protein digestion products.

Absorption of Free Amino Acids

The small intestine has a high capacity to absorb free amino acids. Most L-amino acids can be transported across the epithelium against a concentration gradient, although the need for concentrative transport in vivo is not obvious, since luminal concentrations are usually higher than the plasma levels of 0.1–0.2 mM. Amino acid transport in the small intestine has all the characteristics of carrier-mediated transport, such as discrimination between D- and L-amino acids and energy and temperature dependence. Additionally, genetic defects are known to occur in humans (Clin. Corr. 24.1).

On the basis of genetics and transport experiments, at least five different transport systems for the uptake of L-amino acids from the luminal solution can be distinguished:

1. For neutral amino acids (Ala, Val, Leu, Met, Phe, Tyr, Ile)
2. For basic amino acids and cystine (Lys, Arg, Cys-Cys)
3. For imino acids and glycine (Pro, Hyp, Gly)
4. For acidic amino acids (Asp, Glu)
5. For β-amino acids (β-Ala, taurine)

The mechanisms for concentrative transepithelial transport of L-amino acids appear to be similar to those discussed for D-glucose (see

Figure 24.16). Na^+-dependent transport systems have been identified in the luminal (brush border) membrane and Na^+-independent ones in the contraluminal plasma membrane of small intestinal epithelial cells. Similarly, as for active glucose transport, the energy for concentrative amino acid transport appears to be derived directly from the electrochemical Na^+ gradient and only indirectly from ATP. The amino acids are not chemically modified during membrane transport, although they may be metabolized within the cytoplasmic compartment.

For some time it was thought that the surface enzyme γ-glutamyltransferase plays a specific role in the membrane transport of amino acids (See Chapter 12, Section 12.5). The enzyme catalyzes the transfer of the γ-glutamyl moiety of glutathione (γ-glutamylcysteinylglycine) to a free amino acid, thus forming the dipeptides γ-glutamylamino acid plus cysteinylglycine.

$$\gamma\text{-Glutamylcysteinylglycine} + \text{amino acid} \xrightarrow[(\gamma\text{-glutamyltransferase})]{} \gamma\text{-glutamylamino acid} + \text{cysteinylglycine}$$

With the assumptions that (1) the amino acid is derived from the extracellular space, (2) glutathione is extracted from the cytoplasm, and (3) the dipeptides are released into the cytoplasm, vectorial amino acid transport could be explained by the γ-glutamyltransferase reaction. However, a patient with undetectable γ-glutamyltransferase activity has been described. The patient had apparently normal amino acid transport, suggesting that the enzyme has no physiological role in this respect.

Absorption of Intact Proteins

The fetal and neonatal small intestines can absorb intact proteins. The uptake occurs by endocytosis, that is, the internalization of small vesicles of plasma membrane, which contain ingested macromolecules. The process is also termed "pinocytosis" because of the small size of the vesicles. The small intestinal pinocytosis of protein is thought to be important for the transfer of maternal antibodies (γ-globulins) to the offspring, particularly in rodents. The pinocytotic uptake of proteins is not important for nutrition, and its magnitude usually declines after birth. Persistence of low levels of this process beyond the neonatal period may, however, be responsible for absorption of sufficient quantities of macromolecules to induce antibody formation.

24.5 DIGESTION AND ABSORPTION OF CARBOHYDRATES

Digestion of Carbohydrates

Dietary carbohydrates provide a major portion of the daily caloric requirement. (See Table 24.7.) From a digestive point of view, it is

Table 24.7 Dietary Carbohydrates

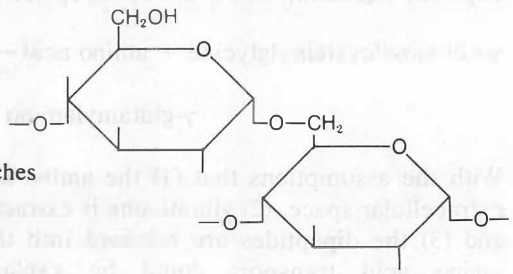
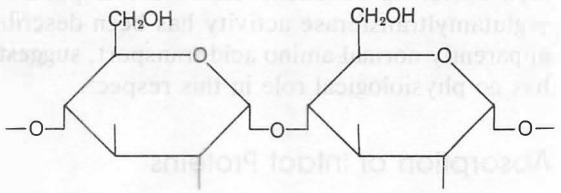
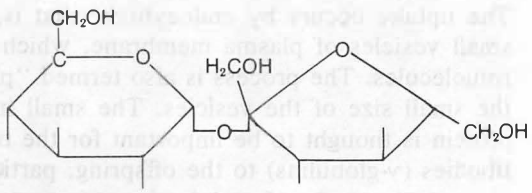
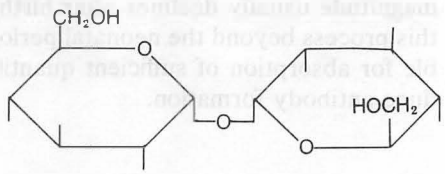
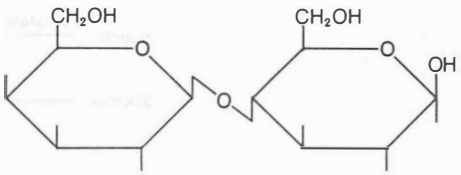
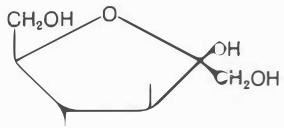
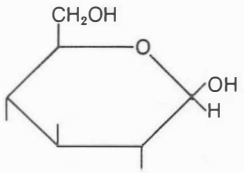
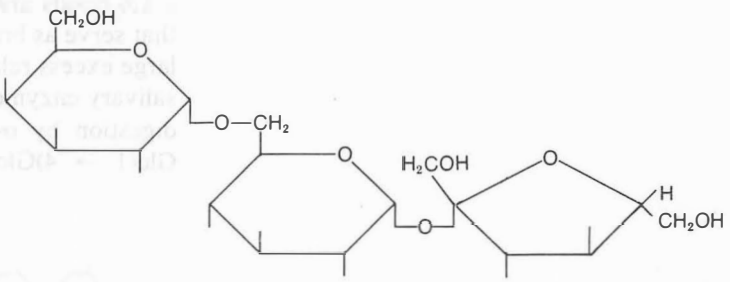
Carbohydrate	*Typical Source*	*Structure*
Amylopectin	Potatoes, rice, corn, bread	α-Glc(1 ⟶ 4)$_n$Glc with α-Glc(1 ⟶ 6) branches 
Amylose	Potatoes, rice, corn, bread	α-Glc(1 ⟶ 4)$_n$Glc 
Sucrose	Table sugar, desserts	α-Glc(1 ⟶ 2)β-Fru 
Trehalose	Young mushrooms	α-Glc(1 ⟶ 1)α-Glc 

Table 24.7 (Continued)

Lactose	Milk, milk products	β-Gal(1⟶4)Glc	
Fructose	Fruit, honey	Fru	
Glucose	Fruit, honey, grape	Glc	
Raffinose	Leguminous seeds	α-Gal(1⟶6)α-Glc(1⟶2)β-Fru	

important to distinguish between mono-, di-, and polysaccharides. Monosaccharides need not be hydrolyzed prior to absorption. Disaccharides require the small intestinal surface enzymes for breakdown into monosaccharides, while polysaccharides depend additionally on pancreatic amylase for degradation. (See Figure 24.22.)

Starch is a major nutrient. It is a plant polysaccharide with a molecular weight of more than 100,000. It consists of either a linear chain of glucose molecules linked by α-1,4-glucosidic bonds

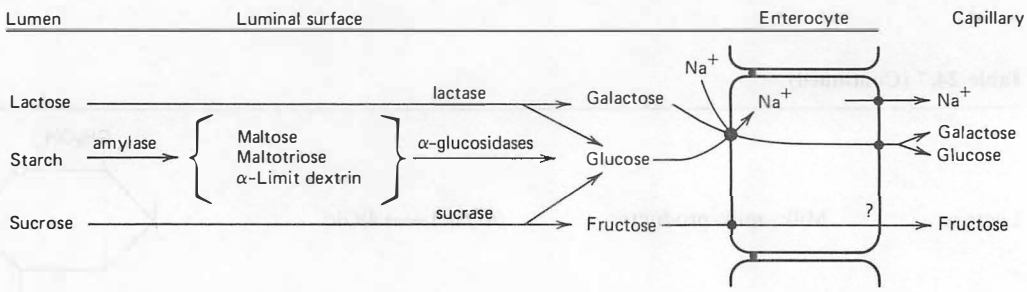


Figure 24.22
Digestion and absorption of carbohydrates.

(amylose) or of a branched chain with branch points made up by α-1,6 linkages (amylopectin). The ratio of 1,4- to 1,6-glucosidic bonds is about 20 : 1. Glycogen is an animal polysaccharide similar in structure to amylopectin. The two compounds differ in terms of the number of branch points, which occur more frequently in glycogen.

Hydrated starch and glycogen are attacked by the endosaccharidase α-amylase, which is present in saliva and pancreatic juice. (See Figure 24.23.) Hydration of the polysaccharides occurs during heating, which, therefore, is essential for efficient digestion of starch. Amylase has specificity for internal α-1,4-glucosidic bonds; α-1,6 bonds are not attacked, nor are α-1,4 bonds of glucose units that serve as branch points. The pancreatic isoenzyme is secreted in large excess relative to starch intake and is more important than the salivary enzyme from a digestive point of view. The products of the digestion by α-amylase are mainly the disaccharide maltose [α-Glc(1 → 4)Glc], the trisaccharide maltotriose [α-Glc(1 → 4)α-

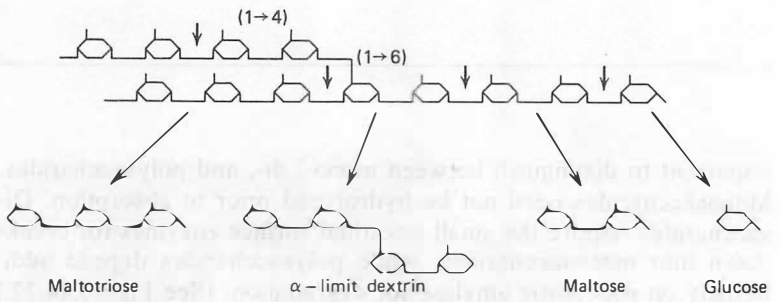


Figure 24.23
Digestion of amylopectin by salivary and pancreatic α-amylase.

Glc(1 → 4)Glc], and so-called α-limit dextrins containing on the average eight glucose units with one or more α-1,6-glucosidic bonds.

Final hydrolysis of di- and oligosaccharides to monosaccharides is carried out by surface enzymes of the small intestinal epithelial cells (Table 24.8). Most of the surface oligosaccharidases are exoenzymes which clip off one monosaccharide at a time from the nonreducing end. The activities of α-glucosidases of the small intestine normally exceed those necessary to complete digestion of starch. Similarly, the capacity to hydrolyze sucrose (table sugar) exceeds that of intake. In contrast, β-galactosidase (lactase) can be rate-limiting in humans for hydrolysis and utilization of lactose, the major milk carbohydrate (Clin. Corr. 24.2).

Di-, oligo-, and polysaccharides that are not hydrolyzed by α-amylase and/or small intestinal surface enzymes cannot be absorbed; therefore they reach the lower tract of the intestine, which from the lower ileum on contains bacteria. Bacteria can utilize many more types of carbohydrates than humans, because their saccharidases are capable of splitting many types of bonds that are not affected by the human enzymes. The monosaccharides that are released as a result of bacterial enzymes are predominantly anaerobically metabolized by the bacteria themselves, resulting in such degrada-

Table 24.8 Di- and Oligosaccharidases of the Luminal Plasma Membrane in the Small Intestine

Enzyme	*Specificity*	*Natural Substrate*	*Product*
exo-1,4-α-Glucosidase (γ-amylase)	α-(1 ⟶ 4)Glucose	Amylose	Glucose
Oligo-1,6-glucosidase (isomaltase)	α-(1 ⟶ 6)Glucose	Isomaltose, α-dextrin	Glucose
α-Glucosidase (maltase)	α-(1 ⟶ 4)Glucose	Maltose, maltotriose	Glucose
Sucrose-α-glucosidase (sucrase)	α-glucose	Sucrose	Glucose, fructose
α,α-Trehalase	α-(1 ⟶ 1)Glucose	Trehalose	Glucose
β-Glucosidase	β-glucose	Glucosyl-ceramide	Glucose, ceramide
β-Galactosidase (lactase)	β-galactose	Lactose	Glucose, galactose

CLIN. CORR. 24.2 DISACCHARIDASE DEFICIENCY

Intestinal disaccharidase deficiencies are encountered relatively frequently in humans. Deficiency can be present in either a single enzyme or several enzymes for a variety of reasons (genetic defect, physiological decline with age, or as result of "injuries" to the mucosa). Of the disaccharidases, lactase is the most common enzyme with an absolute or relative deficiency, which is experienced as milk intolerance. The consequences of an inability to hydrolyze lactose in the upper small intestine are (1) inability to absorb lactose, and (2) bacterial fermentation of ingested lactose in the lower small intestine. Bacterial fermentation results in the production of gas (distension of gut, flatulence) and osmotically active solutes that draw water into the intestinal lumen (diarrhea).

tion products as short-chain fatty acids, lactate, hydrogen gas (H_2), methane (CH_4), and carbon dioxide (CO_2). These compounds can cause fluid secretion, increased intestinal motility, and cramps, either because of increased intraluminal osmotic pressure, distension of the gut, or because of a direct irritating effect of the bacterial degradation products on the intestinal mucosa.

The well-known problem of flatulence after ingestion of leguminous seeds (beans, peas, soya) can be traced to oligosaccharides, which cannot be hydrolyzed by human intestinal enzymes. The leguminous seeds contain modified sucrose to which one or more galactose moieties are linked. The glycosidic bonds of galactose are in the α configuration, which can only be split by bacterial enzymes. The simplest sugar of this family is raffinose [α-Gal(1⟶6)α-Glc (1⟶2) β-Fru] (see Table 24.7).

Trehalose [α-Glc(1⟶1)α-Glc] is a disaccharide that occurs in young mushrooms. The digestion of this sugar requires a special disaccharidase, trehalase.

Absorption of Monosaccharides

The major monosaccharides that result from the digestion of di- and polysaccharide hydrolysis are D-glucose, D-galactose, and D-fructose. Absorption of these and other minor monosaccharides are carrier-mediated processes that exhibit such features as substrate specificity, stereospecificity, saturation kinetics, and inhibition by specific inhibitors.

At least two types of transport systems are known to catalyze the uptake of monosaccharides from the lumen into the cell: (1) a Na^+ monosaccharide cotransport system with high specificity for D-glucose and D-galactose, which catalyzes "active" sugar transport: (2) a Na^+-*independent*, facilitated-diffusion type of monosaccharide transport system with specificity for D-fructose. Additionally, a Na^+-*independent* monosaccharide transport system with specificity for D-glucose and D-galactose is present in the contraluminal plasma membrane. It mediates exit of monosaccharides from epithelial cells (for principles of transepithelial glucose transport (see page 1147). The properties of the two transport systems accepting D-glucose have been investigated in some detail and are compared in Table 24.9. Although overlapping, the substrate specificity and the sensitivity to inhibitors suggest that different proteins are involved in Na^+-*dependent* and Na^+-*independent* glucose transport. The Na^+-glucose cotransport is inhibited by low concentrations of the plant glycoside phlorizin (Figure 24.24), whereas the Na^+-independent one is inhibited by cytochalasin B, (Figure 24.25), which is structurally unrelated to sugars. The inhibitors are not

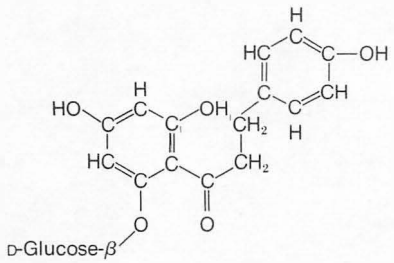


Figure 24.24
Phlorizin (phloretin-2′-β-glucoside).

Table 24.9 Characteristics of Glucose Transport Systems in the Plasma Membranes of Enterocytes

Characteristic	*Luminal*	*Contraluminal*
Effect of Na^+	Cotransport with Na^+	None
Good substrates	D-Glc, D-Gal, α-methyl-D-Glc	D-Glc, D-Gal, D-Man, 2-deoxy-D-Glc
Inhibition by	Phlorizin	Cytochalasin B

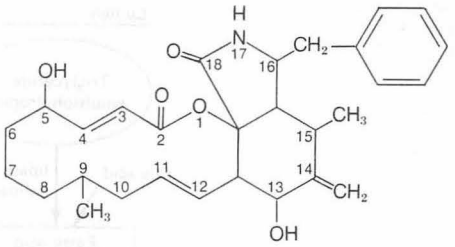


Figure 24.25
Cytochalasin B.

specific for the intestine, but are apparently effective throughout the body, that is, phlorizin also in the Na^+-*dependent* glucose transport of renal proximal tubules and cytochalasin in the Na^+-*independent* one of erythrocytes, adipose tissue, muscle cells, and other tissues.

24.6 DIGESTION AND ABSORPTION OF LIPIDS

General Considerations

An adult man ingests about 60–100 g of fat per day. Triglycerides constitute more than 90% of the dietary fat. The rest is made up of phospholipids, cholesterol, cholesterol esters, and free fatty acids. Additionally, 1–2 g cholesterol and 4–5 g phosphatidylcholine (lecithin) are secreted into the small intestinal lumen as constituents of bile.

Lipids are defined by their good solubility in organic solvents. Conversely, they are sparingly or not at all soluble in aqueous solutions. The poor water solubility presents problems for digestion because the substrates are not easily accessible to the digestive enzymes in the aqueous phase. Additionally, even if ingested lipids are hydrolyzed into simple constituents, the products tend to aggregate to larger complexes that make poor contact with the cell surface and therefore are not easily absorbed. These problems are overcome by (1) increases in the interfacial area between the aqueous and the lipid phase, and (2) "solubilization" of the hydrolysis products with detergents. Thus changes in the physical state of lipids are intimately connected to chemical changes during digestion and absorption. (See Figure 24.26.)

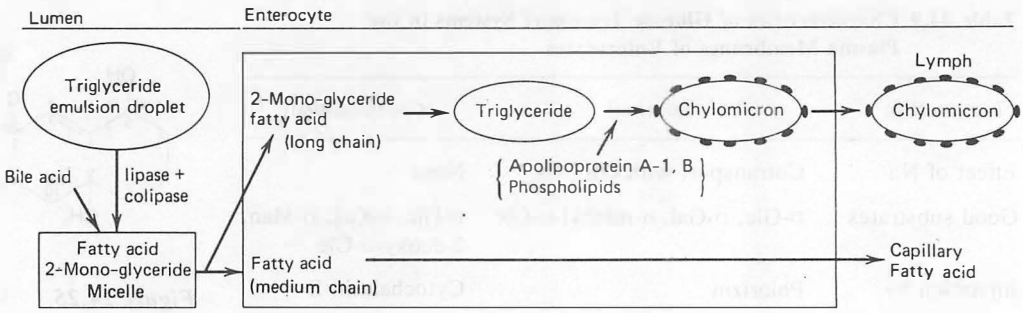


Figure 24.26
Digestion and absorption of lipids.

At least five different phases can be distinguished:

1. Hydrolysis of triglycerides to free fatty acids and monoglycerides
2. Solubilization of free fatty acids and monoglycerides by detergents (bile acids) and transportation from the intestinal lumen toward the cell surface
3. Uptake of free fatty acids and monoglycerides into the cell and resynthesis to triglycerides
4. Packaging of newly synthesized triglycerides into special lipid globules, called chylomicrons
5. Exocytosis of chylomicrons from cells and release into lymph

Digestion of Lipids

The digestion of lipids is initiated in the stomach by an acid-stable lipase. However, the rate of hydrolysis is slow because the ingested triglycerides form a separate lipid phase with a limited water-lipid interface. The lipase adsorbs to that interface and converts triglycerides into fatty acids and monoglycerides. (See Figure 24.27.) The importance of the initial hydrolysis is that some of the water-immiscible triglycerides are converted to products that possess both polar and nonpolar groups. Such products spontaneously adsorb to water–lipid interfaces, and therefore are said to be *surfactive.* In effect, surfactants confer a hydrophilic surface to lipid droplets and thereby provide a stable interface with the aqueous environment. Among the dietary lipids, free fatty acids, monoglycerides, and phospholipids are the major surfactants. One of the effects of the action of lipase is then a release of surfactive molecules and through these an increase in interfacial area between lipid and water. At

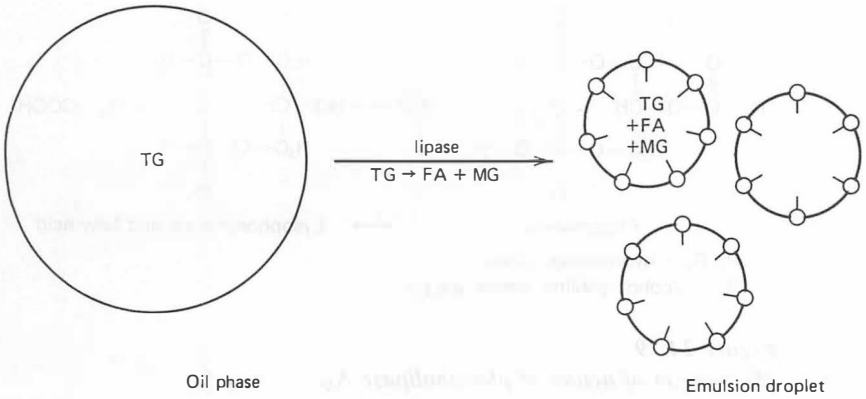


Figure 24.27
Changes in physical state during triglyceride digestion.
Abbreviations: TG = triglyceride, FA = fatty acid, MG = monoglyceride.

constant volume of the lipid phase, any increase in interfacial area produces dispersion of the lipid phase into smaller droplets (emulsification). Thus gastric lipase autocatalytically enhances the availability of more triglyceride substrate through an increase in interfacial area.

The major enzyme for triglyceride hydrolysis is the pancreatic lipase (Figure 24.28). This enzyme is specific for esters in the α position of glycerol and prefers long-chain fatty acids of more than 10 carbon atoms. Triglyceride hydrolysis by the pancreatic enzyme also occurs at the water–lipid interface of emulsion droplets. The products are free fatty acids and β-monoglycerides. The purified form of the enzyme is strongly inhibited by the bile acids that normally are present in the small intestine during lipid digestion. The

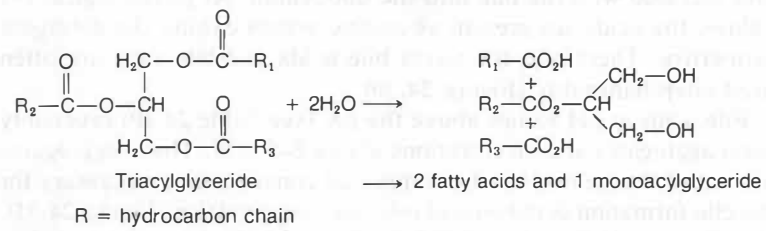


Figure 24.28
Mechanism of action of lipase.

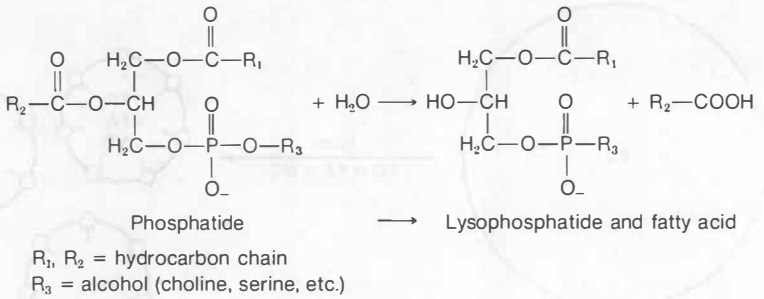


Figure 24.29
Mechanism of action of phospholipase A_2.

problem of inhibition is overcome by the addition of a protein of molecular weight of about 11,000, which is normally present in pancreatic juice. This protein forms a stabilizing complex with lipase, which is no longer sensitive to bile acid inhibition. Because the protein aids lipase in its activity, it is called colipase.

In addition to lipase, pancreatic juice contains another, less specific lipid esterase. This enzyme acts on cholesterol esters, monoglycerides, or other lipid esters, such as esters of vitamin A with carboxylic acids. In contrast to triglyceride lipase, the less specific lipid esterase requires bile acids for activity.

Phospholipids are broken down by specific phospholipases. Pancreatic secretions are especially rich in the proenzyme for phospholipase A_2 (Figure 24.29). As other pancreatic proenzymes, this one, too, is activated by trypsin. Phospholipase A requires bile acids for activity.

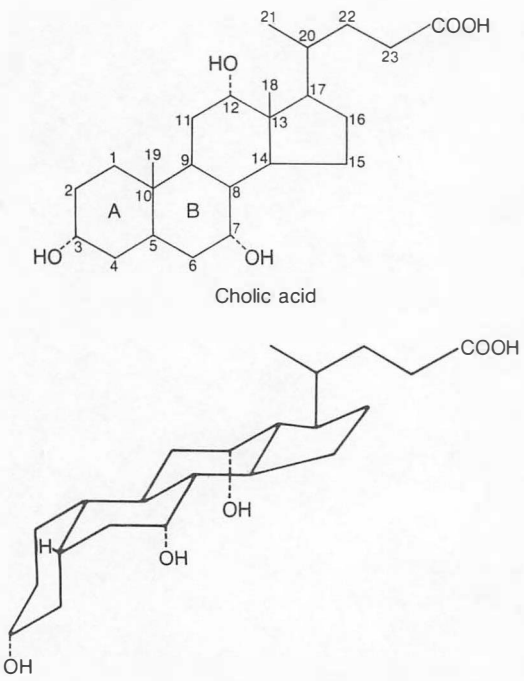


Figure 24.30
Cholic acid, a bile acid.

Role of Bile Acids in Lipid Absorption

Bile acids are biological detergents that are synthesized by the liver and secreted with the bile into the duodenum. At physiological pH values, the acids are present as anions, which exhibit the detergent properties. Therefore, the terms bile acids and bile salts are often used interchangeably (Figure 24.30).

Bile acids at pH values above the pK (see Table 24.10) reversibly form aggregates at concentrations above 2–5 mM. These aggregates are called "micelles," and the minimal concentration necessary for micelle formation is the *critical micellar concentration* (Figure 24.31). Micelles are distinct from lipid emulsion droplets because micelles are smaller than emulsion droplets, and bile acid in the micelle is in equilibrium with free bile acid in solution.

The arrangements of bile acids in micelles is such that the hydrophobic portions of the molecule are removed from contact with water, while hydrophilic groups remain exposed to the water molecules. The hydrophobic region of bile acids is formed by one surface of the fused ring system, while the carboxylate or sulfonate ion and the hydroxyl groups on the other side of the ring system are hydrophilic. Since the major driving forces for micelle formation are the removal of apolar, hydrophobic groups from and the interaction of polar groups with water molecules, the distribution of polar and apolar regions places some constraints on the stereochemical arrangements of bile acid molecules within a micelle. Four bile acid molecules are sufficient to form a very simple micelle as shown in Figure 24.32.

Bile salt micelles can solubilize other lipids, such as phospholipids and fatty acids. These mixed micelles have disklike shapes whereby the phospholipids and fatty acids form a bilayer and the bile acids occupy the edge positions, rendering the edge of the disk hydrophilic (Figure 24.33). Within the mixed phospholipid-bile acid micelles, other water-insoluble lipids, such as cholesterol, can be accommodated and thereby "solubilized" (for potential problems see Clin. Corr. 24.3).

During triglyceride digestion, free fatty acids and monoglycerides are released at the surface of fat emulsion droplets. In contrast to triglycerides, which are water-insoluble, free fatty acids and monoglycerides are slightly water-soluble, and molecules at the surface equilibrate with those in solution. The latter in turn become incorporated into bile acid micelles. Thus the products of triglyceride hydrolysis are continuously transferred from emulsion droplets to the micelles.

Micelles provide the major vehicle for moving lipids from the intestinal lumen to the cell surface where absorption occurs (Figure 24.34). Because the fluid layer next to the cell surface is poorly mixed, the major transport mechanism for solute flux across this "unstirred" fluid layer is diffusion down the concentration gradient. With this type of transport mechanism, the delivery rate of nutrients at the cell surface is proportional to their concentration difference between luminal bulk phase and cell surface. Obviously, the unstirred fluid layer presents problems for sparingly soluble or insoluble nutrients, in that reasonable delivery rates cannot be achieved. Bile acid micelles overcome this problem for lipids by increasing their effective concentration in the unstirred layer. (See Figure 24.34.) The increase in transport rate is nearly proportional to the increase in effective concentration and can be 1,000-fold over that of individually solubilized fatty acids, in accordance with the different solubility of fatty acids as micelles or as individual molecules. This relationship

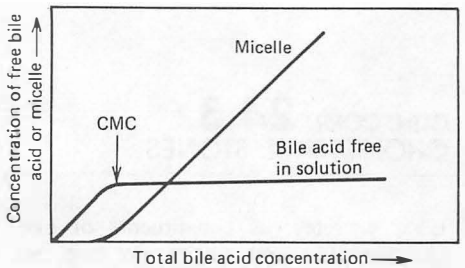


Figure 24.31
Solubility properties of bile acids in aqueous solutions.
Abbreviation: CMC = critical micellar concentration.

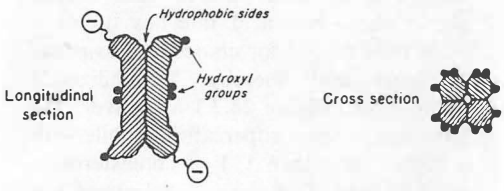


Figure 24.32
Diagrammatic representation of the Na^+ cholate micelle.
Reproduced with permission from D. M. Small, Biochim. Biophys. Acta 176:178, 1969.

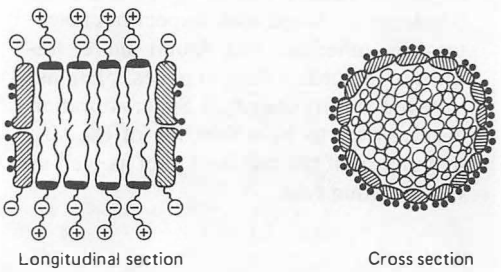


Figure 24.33
Diagrammatic representation of the mixed lecithin–bile acid micelle.
Reproduced with permission from D. M. Small, Biochim. Biophys. Acta 176:178, 1969.

CLIN. CORR. 24.3 CHOLESTEROL STONES

Liver secretes, as constituents of bile, phospholipids and cholesterol together with bile acids. Because of the limited solubility of cholesterol, its secretion in bile can result in cholesterol stone formation in the gallbladder. Stone formation is a relatively frequent complication of bile secretion in humans.

Cholesterol molecules are practically insoluble in aqueous solutions. However, they can be incorporated into mixed phospholipid–bile acid micelles up to a mole ratio of 1:1 for cholesterol to phospholipids and thereby "solubilized." (Clin. Corr. Figure 24.3.) Moreover, the liver can produce supersaturated bile with a higher ratio than 1:1 of cholesterol to phospholipid. This excess cholesterol has a tendency to crystallize out. Such bile with excess cholesterol is considered *lithogenic*, that is, stoneforming. The crystal formation usually occurs in the gallbladder, rather than the hepatic bile ducts, because contact times between bile and any crystallization nuclei are greater in the gallbladder. The tendency to secrete bile supersaturated with respect to cholesterol is inherited and found more frequently in females than in males, often associated with obesity. Supersaturation also appears to be a function of the size and nature of the bile acid pool as well as the secretion rate.

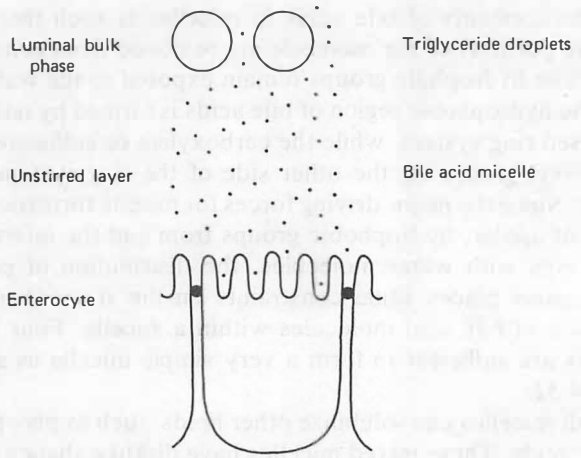


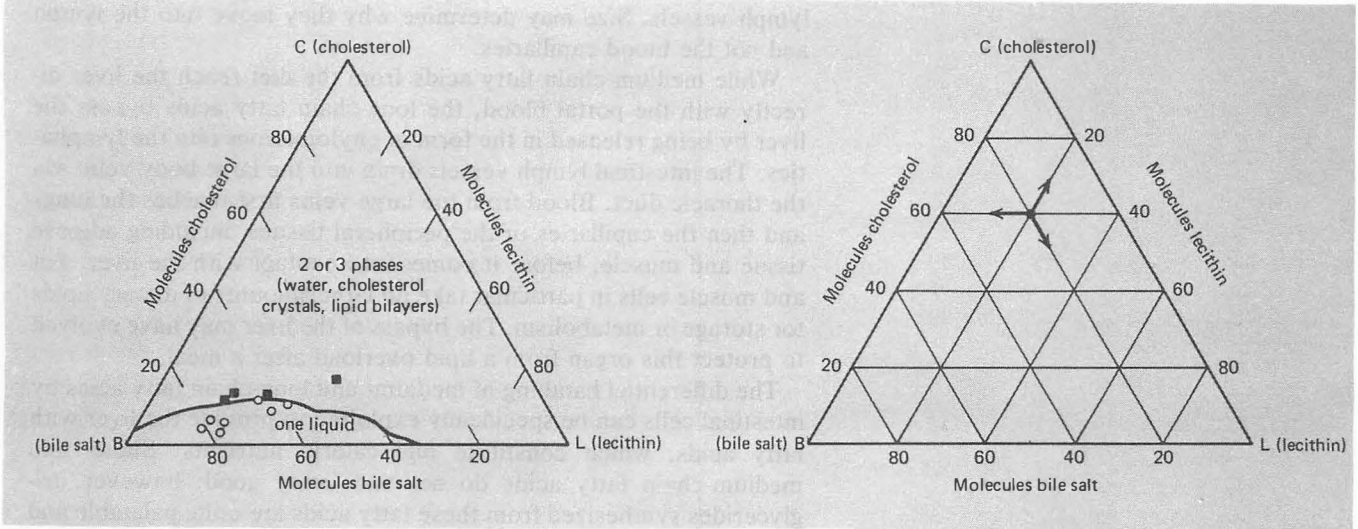
Figure 24.34
Role of bile acids in lipid absorption.

between flux and effective concentration holds because the diffusion constant, the second parameter that determines the flux, is only slightly smaller for the mixed micelles as compared to micellar constituents free in solution. Thus, efficient lipid absorption depends on the presence of sufficient bile acids to "solubilize" the ingested and hydrolyzed lipids in micelles. In the absence of bile acids, the absorption of triglycerides does not completely stop, although the efficiency is drastically reduced.The residual absorption depends on the slight water solubility of the free fatty acids and monoglycerides. Unabsorbed lipids reach the lower intestine where a small part can be metabolized by bacteria. The bulk of unabsorbed lipids, however, is excreted with the stool (=steatorrhea).

Micelles also serve as transport vehicles through the unstirred fluid layers for those lipids that are even less water-soluble than fatty acids, such as cholesterol or the vitamins A and K. For these compounds, bile acid secretion is absolutely essential for absorption by the intestine.

Absorption of Lipids

The uptake of lipids by the epithelial cells of the small intestine can be explained on the basis of diffusion through the plasma membrane. Absorption is virtually complete for fatty acids and monoglycerides, which are slightly water-soluble. It is less efficient for water



Clinical Correlation Figure 24.3
Diagram of the physical states of mixtures of 90% water and 10% lipid.
The 10% lipid are made up of bile acids, lethicin, and cholesterol, and the triangle represents all possible ratios of the three lipid constituents. Each point within the triangle corresponds to a particular composition of the three components, which can be read off the graph as indicated; each point on one of the sides corresponds to a particular composition of just two components. The left triangle contains the composition of gallbladder bile samples from patients without stones (○) *and with cholesterol stones* (■).

Reproduced with permission from A. F. Hofmann and D. M. Small, Annu. Rev. Med. 18:362, 1967. Copyright 1967 by Annual Reviews Inc.

insoluble lipids. For example, only 30–40% of the dietary cholesterol is absorbed.

Within the intestinal cell, the fate of absorbed fatty acids depends on chain length. Fatty acids of medium chain length (6–10 carbon atoms) pass through the cell into the portal blood without modification. In contrast, the long-chain fatty acids (>12 carbon atoms) are resynthesized into triglycerides within the endoplasmic reticulum. Glycerol for this process is provided by the absorbed 2-monoglycerides, and any excess free fatty acids depend for esterification on glycerol derived from glucose. The resynthesized triglycerides form large lipid globules within the cisternae of the endoplasmic reticulum. To keep them suspended in the aqueous phase, phospholipids and special proteins, termed apolipoproteins, are added as surfactants. The intestinal apolipoproteins are different from those of the liver and are classified as A-1 and B. The lipid globules can reach sizes of up to several microns (micrometers). Because of their size and their release into lymph, the globules are called "chylomicrons" (chyle = milky lymph that is present in the intestinal lymph vessels, lacteals, and the thoracic duct after a lipid meal; the word chyle is derived from the Greek *chylos,* which means juice). The chylomicrons are released from the cells by exocytosis into the intercellular space, that is, the chylomicrons move from the endoplasmic reticulum to the contraluminal cell surface in membrane-bound vesicles, which then fuse with the plasma membrane. From the intercellular space, chylomicrons then pass into the

lymph vessels. Size may determine why they move into the lymph and not the blood capillaries.

While medium-chain fatty acids from the diet reach the liver directly with the portal blood, the long-chain fatty acids bypass the liver by being released in the form of chylomicrons into the lymphatics. The intestinal lymph vessels drain into the large body veins via the thoracic duct. Blood from the large veins first reaches the lungs and then the capillaries of the peripheral tissues, including adipose tissue and muscle, before it comes into contact with the liver. Fat and muscle cells in particular take up large amounts of dietary lipids for storage or metabolism. The bypass of the liver may have evolved to protect this organ from a lipid overload after a meal.

The differential handling of medium- and long-chain fatty acids by intestinal cells can be specifically exploited to provide the liver with fatty acids, which constitute high-caloric nutrients. Short- and medium-chain fatty acids do not taste very good; however, triglycerides synthesized from these fatty acids are quite palatable and can be used as part of the diet. In the small intestine the triglycerides would be hydrolyzed by pancreatic lipase and thus provide fatty acids that reach and can be utilized by the liver.

24.7 BILE ACID METABOLISM

All bile acids are synthesized initially within the liver from cholesterol, but they can be modified by bacterial enzymes during passage through the intestinal lumen. The primary bile acids synthesized by the liver are cholic and chenodeoxycholic acid. The secondary bile acids are derived from the primary bile acids by bacterial dehydroxylation in position 7 of the ring structure, resulting in deoxycholate and lithocholate, respectively (Figure 24.35).

Primary and secondary bile acids are reabsorbed by the intestine into the portal blood, taken up by the liver, and then resecreted with bile. Within the liver, primary as well as secondary bile acids are linked to either glycine or taurine via a peptide bond. These derivatives are called glyco- and tauro- conjugates, respectively, and constitute the forms that are secreted into bile. With the conjugation, the carboxyl group of the unconjugated acid is replaced by an even more polar group. The pK values of the carboxyl group of glycine and of the sulfonyl group of taurine are lower than that of unconjugated bile acids, so that conjugated bile acids remain ionized over a wider pH range (Table 24.10). The conjugation is partially reversed within the intestinal lumen by hydrolysis of the peptide bond.

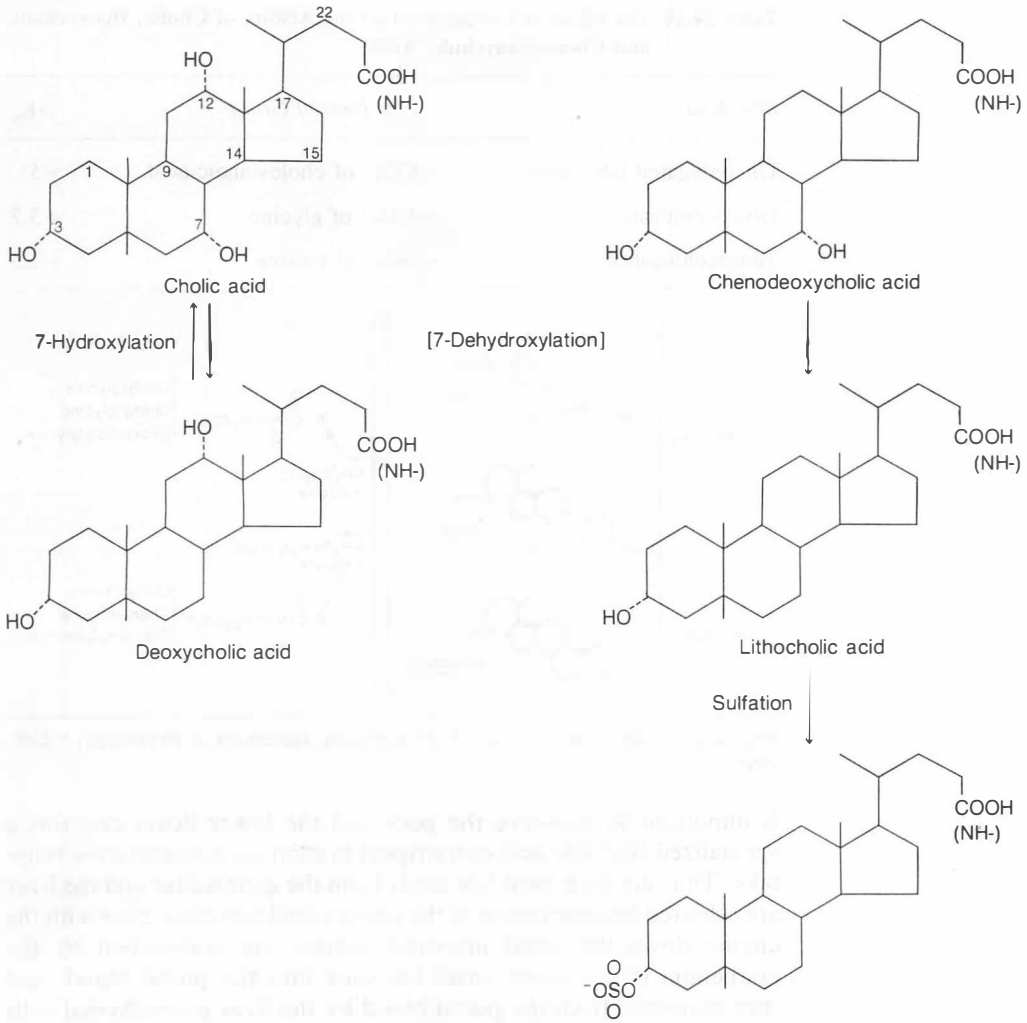
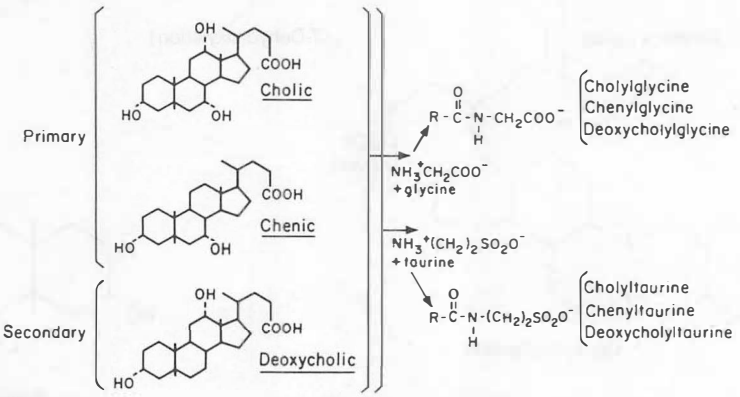


Figure 24.35
Bile acid metabolism in the rat.
Straight arrows indicate reactions catalyzed by liver enzymes; bent arrows indicate those of bacterial enzymes within the intestinal lumen. (NH–), glycine or taurine conjugate of the bile acids.

The total amount of conjugated and unconjugated bile acids secreted per day by the liver is 16–70 g for an adult. As the total body pool is only 3–4 g, bile acids have to recirculate 5–14 times each day between the intestinal lumen and the liver. Reabsorption of bile acids

Table 24.10 The Effect of Conjugation on the Acidity of Cholic, Deoxycholic, and Chenodeoxycholic Acid

Bile Acid	*Ionized Group*	pK_a
Unconjugated bile acids	$-CO_2^-$ of cholestanoic acid	≈5
Glycoconjugates	$-CO_2^-$ of glycine	≈3.7
Tauroconjugates	$-SO_3^-$ of taurine	≈1.5



Reproduced with permission from A. F. Hofmann, Handbook of Physiology, 5:2508, 1968.

is important to conserve the pool and the lower ileum contains a specialized Na^+-bile acid cotransport system for concentrative reuptake. Thus during a meal bile acids from the gallbladder and the liver are released into the lumen of the upper small intestine, pass with the chyme down the small intestinal lumen, are reabsorbed by the epithelium of the lower small intestine into the portal blood, and then extracted from the portal blood by the liver parenchymal cells during passage through the liver. The process of secretion and reuptake is referred to as the enterohepatic circulation (Figure 24.36). The reabsorption of bile acids by the intestine is quite efficient as only about 0.5 g bile acids escape reuptake each day and are secreted with the feces. Serum levels of bile acids normally vary with the rate of reabsorption and therefore are highest during a meal when the enterohepatic circulation is most active.

Cholate, deoxycholate, chenodeoxycholate, and their conjugates continuously participate in the enterohepatic circulation. In contrast, most of the lithocholic acid that is produced by the action of bacterial enzymes within the intestine is sulfated during the next passage through the liver. The sulfate ester of lithocholic acid is not a

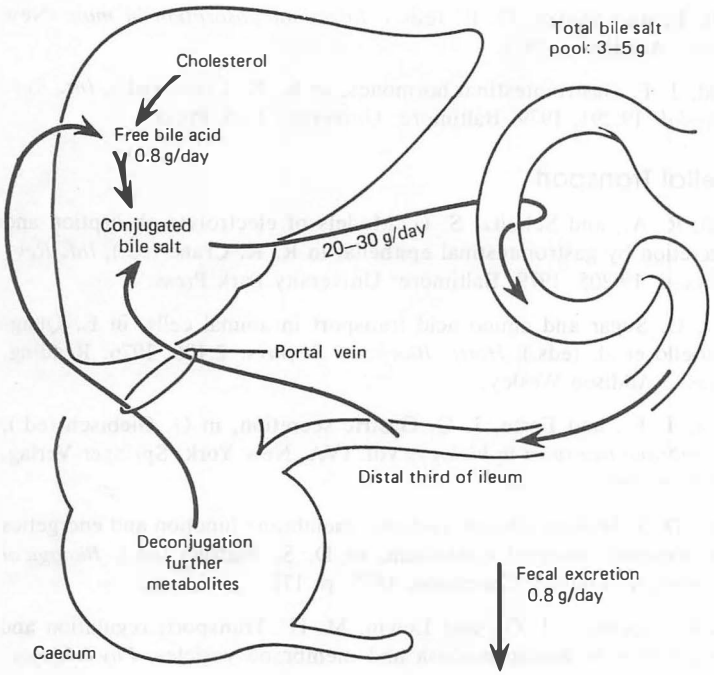


Figure 24.36.
Enterohepatic circulation of bile acids.
Reproduced with permission from M. L. Clark and J. T. Harries in Intestinal Absorption in Man, I. McColl and G. E. Sladen, eds., Academic Press, New York, 1975, p. 195.

substrate for the bile acid transport system in the ileum, and therefore is excreted with the feces.

BIBLIOGRAPHY

General

Gardner, J. D. Regulation of pancreatic exocrine function in vitro: initial steps in the action of secretagogues. *Annu. Rev. Physiol.* 41:35, 1979.

Jamieson, J. D. Membranes and secretions, in G. Weissman and R. Claiborne (eds.), *Cell membranes*. New York: Hospital Practice Publishing Co., 1975.

McColl, I., and Sladen, G. E. (eds.), *Intestinal absorption in man*. New York: Academic, 1975.

Rehfeld, J. F. Gastrointestinal hormones, in R. K. Crane (ed.), *Int. Rev. Physiol.* 19:291, 1979. Baltimore: University Park Press.

Epithelial Transport

Frizzell, R. A., and Schultz, S. G. Models of electrolyte absorption and secretion by gastrointestinal epithelia, in R. K. Crane (ed.), *Int. Rev. Physiol.* 19:205, 1979. Baltimore: University Park Press.

Hopfer, U. Sugar and amino acid transport in animal cells, in E. Quagliariello et al. (eds.), *Horiz. Biochem. Biophys.* 2:106, 1976. Reading, Mass.: Addison-Wesley.

Machen, T. E., and Forte, J. G. Gastric secretion, in G. Giebisch (ed.), *Membrane transport in biology*, vol. IVA. New York: Springer-Verlag, 1979, p. 693.

Parsons, D. S. Multimembrane systems: membrane function and energetics in intestinal mucosal epithelium, in D. S. Parsons (ed.), *Biological membrane*. Oxford: Clarendon, 1975, p. 172.

Sachs, G., Spenney, J. G., and Lewin, M. H^+ Transport: regulation and mechanism in gastric mucosa and membrane vesicles. *Physiol. Rev.* 58:106, 1978.

Schulz, I., and Ullrich, K. J. Transport processes in the exocrine pancreas, in G. Giebisch (ed.), *Membrane transport in biology*, vol. IVA. New York: Springer-Verlag, 1979, p. 811.

Digestion and Absorption of Proteins

Freeman, H. J., and Kim, Y. S. Digestion and absorption of proteins. *Annu. Rev. Med.* 29:99, 1978.

Silk, D. B. A., and Dawson, A. M. Intestinal absorption of carbohydrate and protein in man, in R. K. Crane (ed.), *Int. Rev. Physiol.* 19:151, 1979. Baltimore: University Park Press.

Sleisenger, M. H., and Kim, Y. S. Protein digestion and absorption. *N. Engl. J. Med.* 300:659, 1979.

Digestion and Absorption of Carbohydrates

Cristofaro, E., Mottu, F., and Wuhrmann, J. J. Involvement of the raffinose family of oligosaccharides in flatulence, in H. L. Sipple and K. W. McNutt (eds.), *Sugars in nutrition*. New York: Academic, 1974, p. 314.

Dahlqvist, A. Enzyme deficiency and malabsorption of carbohydrates, in H. L. Sipple and K. W. McNutt (eds.), *Sugars in nutrition*. New York: Academic, 1974, p. 189.

Gray, G. M. Intestinal digestion and maldigestion of dietary carbohydrates. *Annu. Rev. Med.* 22:391, 1971.

Gray, G. M. Carbohydrate digestion and absorption: role of the small intestines. *N. Engl. J. Med.* 292:1225, 1975.

Digestion and Absorption of Lipids

Rommel, K., and Bohmer, R. (eds.). *Lipid absorption: biochemical and clinical aspects.* Baltimore: University Park Press, 1976.

Bile and Metabolism

Bjoerkhem, I., and Danielsson, H. Biosynthesis and metabolism of bile acids in man, in *Progress in liver diseases*, vol. 5. New York: Grune and Stratton, 1976, p. 215.

Brandt, L. J., and Bernstein, L. H. Bile salts: their role in cholesterol synthesis, secretion, and lithogenesis. *Am. J. Gastroenterol.* 65:17, 1976.

Hofmann, A. F., and Small, D. M. Detergent properties of bile salts: correlation with physiological function. *Annu. Rev. Med.* 18:333, 1967.

25

Principles of Nutrition I

STEPHEN G. CHANEY

MACRONUTRIENTS

25.1 OVERVIEW

Nutrition is best defined as the utilization of foods by living organisms. Since the process of food utilization is clearly biochemical in nature, the major thrust of the next two chapters is a discussion of basic nutritional concepts in biochemical terms. However, simply understanding basic nutritional concepts is no longer sufficient. Nutrition appears to attract more than its share of controversy in our society, and a thorough understanding of nutrition almost demands an understanding of the issues behind these controversies. Thus these chapters also explore the biochemical basis for some of the most important nutritional controversies.

Why so much controversy in the first place? Part of the problem is simply that nonscientists feel competent to be "experts." After all, nutrition is concerned with food, and everyone knows about food. However, part of the problem is scientific. Both food itself and our utilization of it are very complex. Quite often the ideal scientific approach of examining only one variable at a time—isolated from other, related variables—is simply inadequate to handle this complex situation. Human variability raises further problems. In studying biochemistry the tendency is to study metabolic pathways and control systems as if they were universal, yet probably no two people utilize nutrients in exactly the same manner. Finally, it is important to realize that much of our knowledge of nutrient requirements and functions comes from animal studies. The days of using prison "volunteers" for nutritional experimentation are behind us. Yet animals are seldom completely adequate models for human beings, since their biochemical makeup almost always differs (the ability of most other animals to synthesize their own ascorbic acid is just one illustration). Thus, despite the best efforts of many reputable scientists, some of the most important nutritional questions have yet to be answered.

The study of human nutrition can logically be divided into three areas: undernutrition, overnutrition, and ideal nutrition. With respect to undernutrition, the primary concern in this country is not with the nutritional deficiency diseases, which are now quite rare. Today more attention is directed toward potential, or subclinical, nutritional deficiencies. There are three factors that have the potential for causing nutritional deficiencies: poor diet, increased nutrient requirements for growth and development, and increased nutrient requirements due to metabolic stress. Only when two or three components overlap in the same individual (Figure 25.1) do the risks of symptomatic deficiencies become significant. Overnutrition, on the other hand, is a particularly serious problem in developed countries.

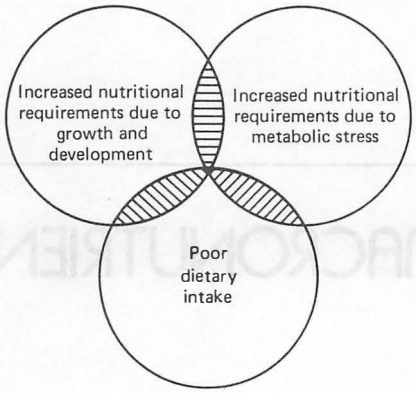


Figure 25.1
Factors affecting individual nutritional status. *Schematic representation of three important risk factors in determining nutritional status. An individual in the white area would have very low risk of any nutritional deficiency, while individuals in the cross-hatched areas would be much more likely to experience some symptoms of nutritional deficiencies.*

Current estimates suggest that between 15 and 30% of the United States population are obese, and obesity is known to have a number of serious health consequences. Finally, there is increasing interest today in the concept of ideal, or optimal nutrition. This is a concept that has meaning only in an affluent society. Only when the food supply becomes abundant enough so that deficiency diseases are a rarity does it become possible to consider the long-range effects of nutrients on health. This is probably the most exciting and least-understood area of nutrition today.

25.2 ENERGY METABOLISM

By now you should be well acquainted with the energy requirements of the body. Much of the foods we eat are converted to ATP and other high energy compounds, which are in turn utilized by the body to drive biosynthetic pathways, generate nerve impulses, and power muscle contraction. We generally describe the energy content of foods in terms of calories. Technically speaking, we are actually referring to the kilocalories of heat energy released by combustion of that food in the body. Some nutritionists today prefer to use the term kilojoule (a measure of mechanical energy), but since the American public is likely to be counting calories rather than joules in the foreseeable future, we will restrict ourselves to that term here. Experimentally, calories can be measured as the heat given off when a food substance is completely burned in a bomb calorimeter, although the caloric value of foods in our body is slightly less due to incomplete digestion and metabolism. The actual caloric values of protein, fat, carbohydrate, and alcohol are roughly 4, 9, 4, and 7 kcal/g, respectively. Given these data and the composition of the food, it is simple to calculate the caloric content (input) of the foods we eat. Calculating caloric content of foods does not appear to be a major problem in this country. Millions of Americans appear to be able to do that with ease. The problem lies in balancing caloric input with caloric output. Where do these calories go?

In practical terms, there are four principal factors that affect individual energy expenditure: surface area (which is related to height and weight), age, sex, and activity levels. (1) The effects of surface area are thought to be simply related to the rate of heat loss by the body—the greater the surface area, the greater the rate of heat loss. While it may seem surprising, a lean individual actually has a greater surface area, and thus a greater energy requirement, than an obese individual of the same weight. (2) Age, on the other hand, may

reflect two factors: growth and lean muscle mass. In infants and children more energy expenditure is required for rapid growth, and this is reflected in a higher basal metabolic rate (rate of energy utilization in the resting state). In adults (even lean adults), muscle tissue is gradually replaced with fat and water during the aging process,

Table 25.1 Recommended Dietary Allowances for Energy Intake[a]

Age (yr) and Sex Group	*Weight*		*Height*		*Energy Needs*	
	kg	*lb*	*cm*	*in*	*kcal*	*range in kcal*
Infants						
0.0–0.5	6	13	60	24	kg × 115	95–145
0.5–1.0	9	20	71	28	kg × 105	80–135
Children						
1–3	13	29	90	35	1,300	900–1,800
4–6	20	44	112	44	1,700	1,300–2,300
7–10	28	62	132	52	2,400	1,650–3,300
Males						
11–14	45	99	157	62	2,700	2,000–3,700
15–18	66	145	176	69	2,800	2,100–3,900
19–22	70	154	177	70	2,900	2,500–3,300
23–50	70	154	178	70	2,700	2,300–3,100
51–75	70	154	178	70	2,400	2,000–2,800
76+	70	154	178	70	2,050	1,650–2,450
Females						
11–14	46	101	157	62	2,200	1,500–3,000
15–18	55	120	163	64	2,100	1,200–3,000
19–22	55	120	163	64	2,100	1,700–2,500
23–50	55	120	163	64	2,000	1,600–2,400
51–75	55	120	163	64	1,800	1,400–2,200
76+	55	120	163	64	1,600	1,200–2,000
Pregnancy					+300	
Lactation					+500	

SOURCE: From Recommended Dietary Allowances, Revised 1980, Food and Nutrition Board, National Academy of Sciences–National Research Council, Washington, D.C.

[a] The data in this table have been assembled from the observed median heights and weights of children, together with desirable weights for adults for mean heights of men (70 in) and women (64 in) between the ages of 18 and 34 years as surveyed in the U.S. population (DHEW/NCHS data).

resulting in a 2% decrease in basal metabolic rate (BMR) per decade of adult life. (3) As for sex, women tend to have a lower BMR than men due to a smaller percentage of lean muscle mass and the effects of female hormones on metabolism. (4) The effect of activity levels on energy requirements is obvious. However, most of us overemphasize the magnitude of this effect. For example, one would need to jog for over an hour to burn up the calories found in one piece of apple pie. The effect of all of these variables on energy requirements can be readily calculated and, assuming light activity levels (a safe assumption for most Americans) and ideal body weight (not such a safe assumption), are presented in tabular form as Recommended Dietary Allowances (RDAs) for energy in Table 25.1.

The RDAs are average values and are widely quoted, but they tell us little about the energy needs of individuals. For example, body composition is known to affect energy requirements, since lean muscle tissue has a higher basal metabolic rate than adipose tissue. This, in part, explains the higher energy requirements of athletes and the lower energy requirements of obese individuals. Hormone levels are important also, since thyroxine, sex hormones, growth hormones, and, to a lesser extent, epinephrine and cortisol are known to increase BMR. The effects of epinephrine and cortisol probably explain in part why severe stress and major trauma significantly increase energy requirements. Fever also causes a significant increase in energy requirements. Finally, energy intake itself has an inverse relationship to expenditure in that during periods of starvation or semistarvation BMR can decrease up to 50%. This is of great survival value in cases of genuine starvation, but not much help to the person who wishes to lose weight on a calorie-restricted diet. Clearly then, the above tables are only general guidelines and may bear little resemblance to the energy needs of a given individial—a fact that is important to remember in the treatment of obesity.

25.3 PROTEIN METABOLISM

Normal Fate of Dietary Protein

Protein carries a certain mystique as a "body-building" food. While it is true that protein is an essential structural component of all cells, protein is equally important for maintaining the output of essential secretions such as digestive enzymes and peptide hormones. Protein is also needed to synthesize the plasma proteins, which are essential for maintaining osmotic balance, transporting substances through

CLIN. CORR. 25.1 VEGETARIAN DIETS AND PROTEIN-ENERGY REQUIREMENTS

One of the most important problems of a purely vegetarian diet (as opposed to a lacto-ovo vegetarian diet), is the difficulty in obtaining sufficient calories and protein. The potential caloric deficit results from the fact that the caloric densities of fruits and vegetables are much less than the meats they replace (30–50 cal/100 g vs 150–300 cal/100 g). The protein problem is generally threefold: (1) most plant products are much lower in protein (1–2 g protein/100 g vs 15–20 g/100 g); (2) most plant protein is of low BV; and (3) some plant proteins are less completely digested. Actually, any reasonably well-designed vegetarian diet will usually provide enough calories and protein for the average adult. In fact, the reduced caloric intake may well be of benefit because strict vegetarians do tend to be lighter than their nonvegetarian counterparts.

However, whereas an adult male may require about 0.8 g protein and 38 cal/kg body weight, a young child may easily require 2–3 times that amount. Similarly a pregnant woman needs an additional 30 g of protein and 300 cal/day and a lactating woman an extra 20 g of protein and 500 cal. Thus both young children and pregnant and lactating women run a risk of protein-energy malnutrition.

However, it is possible to provide sufficient calorie and protein even for these high-risk groups provided the diet is adequately planned. There are three principles that can be followed to design a calorie/protein-sufficient vegetarian diet for young children: (1) whenever possible,

the blood, and maintaining immunity. However, the average adult in this country consumes far more protein than needed to carry out these essential functions. The excess protein is simply treated as a source of energy, with the glucogenic amino acids being converted to glucose and the ketogenic amino acids being converted to fatty acids and keto acids. Both kinds of amino acids will of course eventually be converted to triglyceride in the adipose tissue if fat and carbohydrate supplies are already adequate to meet energy requirements. Thus for most of us the only body-building obtained from high-protein diets is adipose tissue.

It has always been popular to say that the body has no storage depot for protein, and thus adequate dietary protein must be supplied with every meal. However, in actuality, this is not quite accurate. While there is no separate class of "storage" protein, there is a certain percentage of muscle and structural protein that can be considered as nonessential or at least expendable, and which undergoes a constant process of breakdown and resynthesis. In the fasting state the breakdown of this store of body protein is enhanced, and the resulting amino acids are utilized for glucose production, the synthesis of nonprotein nitrogeneous compounds, and for the synthesis of the essential secretory and plasma proteins described above (see also Chapter 14 and Clin. Corr. 25.4). Even in the fed state, some of these amino acids will be utilized for energy production and as biosynthetic precursors. Thus the turnover of body protein is a normal process—and an essential feature of what is called nitrogen balance.

Nitrogen balance (Figure 25.2) is simply a comparison between the intake of nitrogen (chiefly in the form of protein) and the excretion of nitrogen (chiefly in the form of undigested protein in the feces and urea and ammonia in the urine). The normal adult will be in nitrogen equilibrium, with losses just balanced by intake. Negative nitrogen balance can result from an inadequate dietary intake of protein, since the amino acids utilized for energy and biosynthetic reactions are not replaced. However, it also is observed in injury when there is net destruction of tissue and in major trauma or illness in which the body's adaptive response causes increased catabolism of body protein stores (see Chapter 14). A positive nitrogen balance will be observed whenever there is a net increase in the body protein stores, such as in a growing child, a pregnant woman, or a convalescing adult.

Essential Amino Acids

In addition to the amount of protein in the diet, several other factors must be considered. One of these is the complement of essential

include eggs and milk in the diet. They are both excellent sources of calories and high-quality protein. (2) Include liberal amounts of those vegetable foods with high caloric density in the diet. These include nuts, grains, dried beans, and dried fruits. (3) Include liberal amounts of high-protein vegetable foods, at the same meal, which have complementary amino acid patterns.

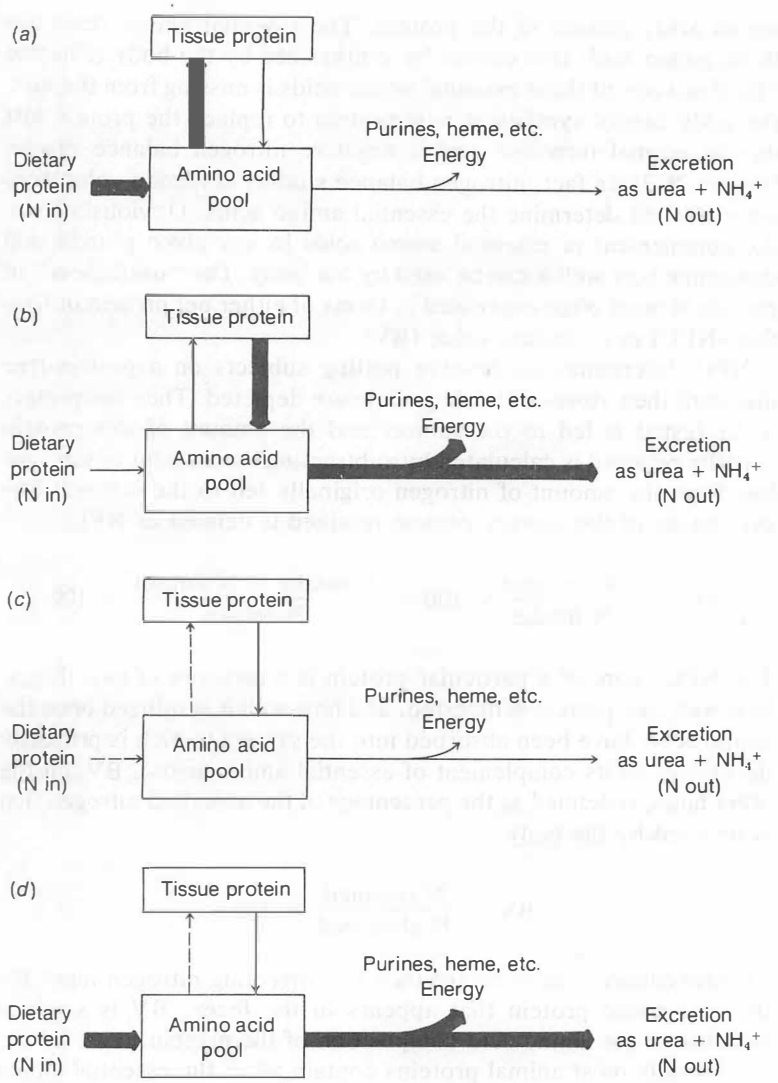


Figure 25.2
Factors affecting nitrogen balance.
Schematic representations of the metabolic interrelationship involved in determining nitrogen balance. (a) Positive nitrogen balance (growth, pregnancy, lactation, recovery from metabolic stress). (b) Negative nitrogen balance (metabolic stress). (c) Negative nitrogen balance (inadequate dietary protein). (d) Negative nitrogen balance (lack of an essential amino acid). Each figure represents the nitrogen balance resulting from a particular set of metabolic conditions. The dominant pathways in each situation are indicated by heavy arrows.

amino acids present in the protein. The essential amino acids are those amino acids that cannot be synthesized by the body (Chapter 12). If just one of these essential amino acids is missing from the diet, the body cannot synthesize new protein to replace the protein lost due to normal turnover, and a negative nitrogen balance results (Figure 25.2). In fact, nitrogen balance studies in human volunteers were used to determine the essential amino acids. Obviously then, the complement of essential amino acids in any given protein will determine how well it can be used by our body. The "usefulness" of protein is most often expressed in terms of either net protein utilization (NPU) or biological value (BV).

NPU determinations involve putting subjects on a protein-free diet until their stores of labile protein are depleted. Then the protein to be tested is fed to the subject and the amount of this protein actually retained is calculated by subtracting the amount of nitrogen lost from the amount of nitrogen originally fed to the subject. The percentage of this dietary protein retained is defined as NPU:

$$\text{NPU} = \frac{\text{N retained}}{\text{N intake}} \times 100 = \frac{\text{N intake} - \text{N output}}{\text{N intake}} \times 100$$

The NPU score of a particular protein is a measure of two things: how well that protein is digested, and how well it is utilized once the amino acids have been absorbed into the system (which is primarily dependent on its complement of essential amino acids). BV, on the other hand, is defined as the percentage of the absorbed nitrogen that is retained by the body:

$$\text{BV} = \frac{\text{N retained}}{\text{N absorbed}} \times 100$$

Experimentally, this is determined by correcting nitrogen input for the undigested protein that appears in the feces. BV is solely a measure of the amino acid composition of the protein.

Generally most animal proteins contain all of the essential amino acids in about the quantities needed by the human body and thus have high NPU scores (see Table 25.2). Vegetable proteins, on the other hand, often lack one or more essential amino acids and may, in some cases, be more difficult to digest. This is reflected by lower NPU scores. Even so, vegetarian diets can provide adequate protein provided (1) enough extra protein is consumed to provide sufficient quantities of the essential amino acids and/or (2) two or more proteins are consumed together, which complement one another in amino acid content. For example, if corn (which is deficient in lysine) is combined with legumes (which are deficient in methionine but rich

Table 25.2 Net Protein Utilization Values of Some Common Foods

Protein	*NPU in Young Children*
Human milk	95
Whole hen egg	87
Cow's milk	81
Polished rice	63
Peanuts	57
Soybean flour	54
Whole wheat	49
Maize	36

SOURCE: Taken from p67 of WHO Technical Report #522.

in lysine), the NPU for the mixture approaches that of animal protein. The adequacy of vegetarian diets with respect to protein and calories is discussed more fully in Clin. Corr. 25.1 and the need for high-quality protein in low-protein renal diets is discussed in Clin. Corr. 25.2.

Effect of Carbohydrate and Fat: Protein Sparing

Another factor that must be considered in determining protein requirements is the dietary intake of fat and carbohydrate. If these components are present in insufficient quantities, some of the dietary protein must be used for energy generation and is unavailable for building and replacing tissue. Thus, as the energy (calorie) content of the diet from carbohydrate and fat increases, the need for protein decreases. This is referred to as protein sparing. Carbohydrate is somewhat more efficient at protein sparing than fat—presumably because carbohydrate can be used as an energy source by almost all tissues, whereas fat cannot. Assuming light activity and consumption of high quality protein, a protein/calorie ratio of 1 : 20 seems adequate, that is, at least 5% of the calories should come from protein.

Normal Adult Protein Requirements

Assuming adequate caloric intake and a NPU of 75% on the mixed protein in the average American diet, the recommended protein intake is 0.8 g/(kg body wt · day). This amounts to about 56 g protein/day for a 70-kg (154-lb) man and about 44 g/day for a 55-kg (120-lb) woman. These recommendations would obviously need to be increased on a vegetarian diet if the overall NPU was less than 75%.

Other Factors Affecting Protein Requirements

Because dietary protein is essential for the synthesis of new body tissue, as well as for maintenance and repair, the need for protein increases markedly during periods of rapid growth. Such growth occurs during pregnancy, infancy, childhood, and adolescence. These recommendations are summarized in Table 25.3.

Once growth requirements have been considered, age does not seem to have much effect on protein requirements. If anything, the protein requirement may slightly decrease with age. However, older people need and generally consume fewer calories, so high-quality protein should provide a larger percentage of their total calories.

Table 25.3 Recommended Dietary Allowances for Protein

Age (yr) and Sex Group	*Weight*		*Height*		*Protein (g)*
	kg	*lb*	*cm*	*in*	
Infants					
0.0–0.5	6	13	60	24	kg × 2.2
0.5–1.0	9	20	71	28	kg × 2.0
Children					
1–3	13	29	90	35	23
4–6	20	44	112	44	30
7–10	28	62	132	52	34
Males					
11–14	45	99	157	62	45
15–18	66	145	176	69	56
19–22	70	154	177	70	56
23–50	70	154	178	70	56
51+	70	154	178	70	56
Females					
11–14	46	101	157	62	46
15–18	55	120	163	64	46
19–22	55	120	163	64	44
23–50	55	120	163	64	44
51+	55	120	163	64	44
Pregnancy					+30
Lactation					+20

SOURCE: From Recommended Dietary Allowances, Revised 1980, Food and Nutrition Board, National Academy of Sciences, National Research Council, Washington, D.C.

CLIN. CORR. 25.2 LOW-PROTEIN DIETS AND RENAL DISEASE

Chronic renal failure is characterized by the buildup of the end products of protein catabolism (mainly urea). Since these toxic end products are responsible for many of the symptoms associated with renal failure, some degree of dietary protein restriction is usually necessary. The amount of protein restriction is dependent on the severity of the disease. It is easy to maintain patients in nitrogen equilibrium for prolonged periods on diets containing as little as 40 g protein/day if the diet is calorically sufficient. However, diets containing less than 40 g/day pose problems. Protein turnover continues and one is forced to walk a tightrope between providing enough protein to avoid negative nitrogen balance, but little enough to avoid buildup of waste products.

The strategy employed in such diets is twofold: (1) to provide a minimum of protein, primarily protein of high BV, and (2) to provide the rest of the daily calories in the form of carbohydrates and fats. The goal is to provide just enough essential amino acids to maintain positive nitrogen balance. In turn, the body should be able to synthesize the nonessential amino acids from other nitrogen-containing metabolites. Enough carbohydrate and fat are provided so that essentially all of the dietary protein can be spared from energy metabolism. With this type of diet, it is possible to maintain a patient on 20 g protein/day for considerable periods.

However, such diets are extremely monotonous and difficult to follow. A typical 20-g protein diet (the Giovannetti diet) is shown below:

Furthermore, some older people may have special protein requirements due to malabsorption problems.

One of the myths of popular nutrition is that athletes need more protein. While there may be a slight increase in protein requirements during periods of intensive body-building, vigorous exercise alone increases only energy needs—not protein requirements. Thus the protein needs of the athlete are similar to those of the nonathlete, provided that the increased caloric requirements are met.

Illness, major trauma, and surgery all cause a major catabolic response by the body. Both energy and protein needs are very large, and the body responds by increasing production of glucagon, glucocorticoids, and epinephrine (see Chapter 14). In these situations the breakdown of body protein is greatly accelerated and a negative nitrogen balance results unless protein intake is increased (Figure 25.2). Although this increased protein requirement is of little significance in short-term illness, it can be vitally important in the recovery of hospitalized patients as discussed in the next section (see also Clin. Corrs. 25.3 and 25.5).

25.4 PROTEIN-ENERGY MALNUTRITION

The most common form of malnutrition in the world today is protein energy malnutrition (PEM). In the developing countries inadequate intake of protein and energy is all too common, and it is usually the infants and young children who suffer most. While the actual symptoms of protein-energy insufficiency vary widely from case to case, it is common to classify all cases as either marasmus or kwashiorkor. Marasmus is usually defined as inadequate intake of both protein and energy and kwashiorkor as inadequate intake of protein in the presence of adequate energy intake. More often than not, the diets associated with marasmus and kwashiorkor may be similar with the kwashiorkor being precipitated by conditions of increased protein demand. The marasmic infant will have a thin, wasted appearance and will be small for his/her age. If the PEM continues long enough the child will be permanently stunted in both physical and mental development. In kwashiorkor, on the other hand, the child will often have a deceptively plump appearance due to edema. Other telltale symptoms associated with kwashiorkor are dry, brittle hair, diarrhea, dermatitis of various forms, and retarded growth. Perhaps the most devastating result of both marasmus and kwashiorkor is

1. One egg *plus* ¾ cup milk or 1 additional egg or 1 oz meat.
2. One-half pound of deglutenized (low protein) wheat bread; all other breads and cereals must be avoided—this includes almost all baked goods.
3. A limited amount of certain low-protein, low-potassium fruits and vegetables.
4. Sugars and fats to fill the rest of the needed calories; however, cakes, pies, and cookies would need to be avoided.

The palatability of this diet would be improved considerably, if it were possible to include more foods containing protein of low BV (vegetables, cereals, bread) for variety, yet still maintain a positive nitrogen balance. One approach to this problem is to use dietary supplements (usually in liquid or tablet form), which provide only the essential amino acids. Another approach is to use supplements containing the carbon skeletons (in the form of α-hydroxy and α-keto acids) of the essential amino acids. Both of these techniques do work, but their feasibility is limited by cost and poor taste.

CLIN. CORR. 25.3 HYPERALIMENTATION AND THE CATABOLIC RESPONSE

The normal metabolic response to infection, trauma, and surgery is a complex and carefully balanced catabolic state. As discussed in the text, epinephrine, glucagon, and cortisol are released, greatly accelerating the rates of lipolysis, proteolysis, and gluconeogenesis. The net result is an increased supply of fatty acids and glucose to meet the increased energy demands of such major stress. The high serum glucose does, in turn, result in some elevation of circulating insulin levels. However, the moderate increase in insulin is more than counterbalanced by the increased levels of epinephrine and other hormones, which render some tissues less responsive to insulin. Skeletal muscle, for example, uses very little of the serum glucose, but continues instead to rely on free fatty acids and its own catabolized protein as a primary source of energy. It also continues to export amino acids (primarily alanine) for use elsewhere in the body, resulting in a very rapid depletion of body protein stores.

For many years, hospitalized patients routinely received intravenous solutions of glucose and water, on the assumption that the primary goal of the catabolic response was to provide adequate glucose levels during periods of stress. The idea was to provide an exogenous source of glucose for increased energy needs and at the same time to raise insulin levels sufficiently to exert a sparing effect on wastage of skeletal protein. In fact, intravenous glucose was very effective at reaching those objectives. Circulating insulin levels do become high enough so that muscle relies primarily on glucose as an energy source, and the overall negative nitrogen balance can be minimized.

However, recent studies have shown that the normal catabolism of skeletal protein during stress situations plays another very important role. It provides the amino acid building blocks for the synthesis of albumin, transferrin, and other essential secretory proteins and for the production of immunocompetent cells and their secretory products. High insulin levels, of course, prevent this response. Patients maintained on 5% glucose while ex-

periencing catabolic stress from surgery, trauma, or infection show a very rapid fall in both serum albumin and immunocompetence and can often develop kwashiorkor-like symptoms in as little as 1–2 weeks. This can delay wound healing and predispose the patient to potentially fatal infections.

Current hospital practice is to use 5% glucose solutions in short-term situations when the patient has been previously well nourished. However, when the patient enters the hospital in a malnourished state or when oral intake is not possible for a long time, protein supplementation is usually included.

CLIN. CORR. 25.4 PROTEIN MODIFIED FASTS

There is at least one advantage to long-term fasting as a means of weight loss. After several days of fasting, appetite is greatly suppressed and the fast can be maintained more easily. The biochemical mechanism for this appetite loss is not really known, but some evidence suggests that it may correlate with increased production of ketone bodies. Starvation, of course, is not the recommended form of weight loss. There are simply too many potential serious medical complications. Among these are low blood pressure, electrolyte imbalances (leading to irregular heart beats), anemia, and hyperuricemia (leading to attacks of gouty arthritis). Most of these changes can be detected and corrected before serious damage can occur. Thus starvation is occasionally used in a carefully controlled hospital situation to produce rapid weight loss in very obese patients.

Unfortunately, even after complete adaptation to starvation, the body is unable to rely solely on its fat stores for energy production. Some protein must also be catabolized, and there does not appear to be any simple way for the body to distinguish between essential protein stores (such as kidney, liver, and cardiac muscle) and nonessential protein stores (skeletal muscle). Thus, prolonged starvation can also lead to irreversible impairment of liver, kidney, and cardiac function.

In recent years several hospitals have been experimenting with so-called protein modified fasts. In this regimen the fast is modified by including 40–100 g of high-quality protein in the diet. This amount of protein does not affect the ketosis or significantly decrease rate of weight loss, but it does allow maintenance of a positive nitrogen balance. Thus, theoretically it should prevent catabolism of essential protein stores. As with unmodified fasting, this approach has been carried out successfully with many patients in a carefully supervised hospital setting. However, it is still considered experimental and potentially quite dangerous if tried in a setting where daily medical supervision is not possible.

Unfortunately, the inherent danger in these diets was forcibly brought to the public's attention when the diet was publicized in the form of a book entitled *Last Chance Diet*. Millions of Americans tried the diet and manufacturers rushed to market a liquid protein made up of hydrolyzed collagen. This fad lasted less than a year before the first deaths were reported. In all, over 40 people were reported to have died—most from heart attacks—before action was initiated to remove liquid protein from the market. Since most of these people were not under close medical supervision during their fast, it is not possible to be certain what caused the deaths.

reduced ability of the afflicted individuals to fight off infection. They have a reduced number of T lymphocytes (and thus diminished cell-mediated immune response) as well as defects in the generation of phagocytic cells and production of immunoglobulins, interferon, and other components of the immune system. Many of these individuals die from secondary infections, rather than from the starvation itself. In the United States, classical marasmus and kwashiorkor are exceedingly rare, but milder forms of protein-energy malnutrition are seen. (See Clin. Corr. 25.4.)

The most common form of PEM seen in the United States today occurs in the hospital setting. A typical course of events is as follows: The patient is not eating well for several weeks or months prior to entering the hospital due to chronic or debilitating illness. He/she enters the hospital with major trauma, severe infection, or for major surgery, all of which cause a large negative nitrogen balance. This is often compounded by difficulties in feeding the patient or by the necessity of fasting the patient in preparation for surgery or diagnostic tests. The net result is PEM as measured by low levels of serum albumin and other serum proteins or by decreased cellular immunity tests. Recent studies have shown that hospitalized patients with demonstrable PEM have delayed wound healing, decreased resistance to infection, increased mortality, and increased length of hospitalization. Currently most major hospitals have instituted programs to monitor the nutritional status of their hospitalized patients and to intervene where necessary to maintain a positive nitrogen and energy balance (Clin. Corr. 25.5).

25.5 EXCESS PROTEIN-ENERGY INTAKE

Much has been said in recent years about the large quantities of protein that the average American consumes. Certainly most of us do consume far more than needed to maintain positive nitrogen balance. The average American currently consumes 99 g of protein, 68% of it from animal sources. However, most studies seem to show that a healthy adult can consume that quantity of protein with no apparent harm. Concern has been raised about the possible effect of high-protein intake on calcium requirements. Some studies suggest that high-protein intakes increase urinary loss of calcium and thus may accelerate the bone demineralization associated with the aging process. However, this issue is far from settled at present.

They could have been due to vitamin or mineral deficiencies or toxic substances in the liquid protein preparations. The deaths may have been caused by an electrolyte imbalance. Low potassium, in particular, would seriously affect heart function. However, the most likely problem was the choice of digested collagen as the protein source. Collagen is of such low BV as to be useless for replenishing vital protein stores.

CLIN. CORR. 25.5 PROVIDING ADEQUATE PROTEIN AND CALORIES FOR THE HOSPITALIZED PATIENT

A highly catabolic hospitalized patient may require 2,500–4,000 cal and 70–100 g protein/day. A patient with severe burns may require even more. The physician has a number of options available to provide this postoperative patient with sufficient calories and protein to insure optimal recovery. When the patient is simply unable to take in enough food, it may be adequate to supplement the diet with high-calorie/high-protein preparations (which are usually mixtures of homogenized corn starch, egg, milk protein, and flavorings). When the patient is unable to take in solid food or unable to digest complex mixtures of foods adequately, elemental diets are usually administered via a nasogastric tube. Elemental diets consist of small peptides or purified amino acids, glucose and dextrins, some fat, vitamins, and electrolytes. These diets are generally low residue and can be used in patients with lower gastrointestinal tract disturbances. They are also very efficiently digested and absorbed in the absence of pancreatic enzymes or

bile salts. These diets are usually sufficient to meet the entire caloric and protein needs of a moderately catabolic patient.

However, when the patient is severely catabolic or unable to digest and absorb foods normally, parenteral (intravenous) nutrition is necessary. The least invasive method is to use a peripheral, slow-flow vein in a manner similar to any other IV infusion. The main limitation of this method is hypertonicity. If the infusion fluid is too hypertonic, there is endothelial cell damage and thrombosis. However, a solution of 5% glucose and 4.25% purified amino acids can safely be used. This solution will usually provide enough protein to maintain positive nitrogen balance, but will rarely provide enough calories for long-term maintenance of a catabolic patient.

The most aggressive nutritional therapy is total parenteral nutrition. Usually an indwelling catheter is inserted into a large fast-flow vessel such as the superior vena cava, so that the very hypertonic infusion fluid can be rapidly diluted. This allows solutions of up to 60% glucose and 4.25% amino acids to be used, providing both sufficient protein and calories for long-term maintenance. All of these methods can be used to prevent or minimize the negative nitrogen balance associated with surgery and trauma. The actual choice of method depends on the patient's condition. As a general rule it is preferrable to use the least invasive technique.

Perhaps the more serious nutritional problem in this country is excessive energy consumption. In fact, obesity is the most frequent nutritional disorder in the United States. A discussion of the treatment of obesity is clearly beyond the scope of this chapter, but it is worthwhile to consider some of the metabolic consequences of obesity. One striking clinical feature of overweight individuals is a marked elevation of free fatty acids, cholesterol, and triglyceride, irrespective of the dietary intake of fat. Why is this? Obesity is obviously associated with an increased number or size of adipose cells. Furthermore, these cells contain fewer insulin receptors and thus respond more poorly to insulin, resulting in increased activity of the hormone-sensitive lipase. The increased lipase activity along with the increased mass of adipose tissue is probably sufficient to explain the increase in circulating free fatty acids. These excess fatty acids are, of course, carried to the liver, where they are repackaged along with cholesterol as very low density lipoprotein particles, leading to higher circulating levels of both triglyceride and cholesterol (for more detail on these metabolic interconversions, see Chapters 10 and 14).

A second striking finding in obese individuals is higher fasting blood sugar levels and decreased glucose tolerance. Fully 80% of adult onset diabetics are overweight. Again the culprit appears to be the decrease in insulin receptors, since many adult onset diabetics have higher than normal insulin levels. Because of these metabolic changes, obesity is one of the primary risk factors in coronary heart disease, hypertension, and diabetes. This is nutritionally significant because all of these metabolic changes are reversible. Quite often reduction to ideal weight is the single most important mode of nutritional therapy. Furthermore, when the individual is at ideal body weight, the composition of the diet becomes a less important consideration in maintaining normal serum lipid and glucose levels.

Any discussion of weight reduction regimens should include a mention of one other metabolic consequence of obesity. Aldosterone levels are also elevated, leading to increased retention of both sodium and water. Furthermore, in some cases aldosterone levels increase even more when the obese person begins dieting. Thus, in effect as the fat stores are metabolized, they are converted to water (which is denser than the fat), and the water may be largely retained. In fact, some individuals may actually observe short-term weight gain on certain diets, even though the diet is working perfectly well in terms of breaking down their adipose tissue. This metabolic fact of life can be devastating psychologically to the dieters, who expect to see quick results for all their sacrifice. This is one major reason for

the popularity of the low-carbohydrate diets, which decrease water retention (Clin. Corr. 25.6).

25.6 CARBOHYDRATES

The chief metabolic role of carbohydrates in the diet is for energy production. Any carbohydrate in excess of that needed for energy is converted to glycogen and triglyceride for long-term storage. The human body can adapt to a wide range of carbohydrate levels in the diet. Diets high in carbohydrate result in higher steady-state levels of glucokinase and some of the enzymes involved in the hexose monophosphate shunt and triglyceride synthesis. Diets low in carbohydrate result in higher steady-state levels of some of the enzymes involved in gluconeogenesis, fatty acid oxidation, and amino acid catabolism. Glycogen stores can also be affected by the carbohydrate content of the diet (Clin. Corr. 25.7). Very low levels of carbohydrate result in a permanent state of ketosis similar to that seen during starvation. The mechanism has been discussed earlier (Chapter 14). If continued over a period of time this ketosis may, in some instances, be detrimental to the patient's health (Clin. Corr. 25.6). Most Americans, of course, do consume more than adequate carbohydrate levels.

The most common nutritional problems involving carbohydrates are seen in those individuals with various carbohydrate intolerances. The most common form of carbohydrate intolerance is diabetes mellitus, which is caused either by lack of insulin production or lack of insulin receptors. This causes an intolerance to glucose and those simple sugars that can be readily converted to glucose. The dietary treatment of diabetes usually involves limiting intake of most simple sugars, and increasing intake of those carbohydrates that are better tolerated, mostly complex carbohydrates, but including some simple sugars such as fructose and sorbitol. Lactase insufficiency is also a common disorder of carbohydrate metabolism affecting over 30 million people in the United States alone. It is most prevalent among blacks, Asians, orientals, and South Americans. Without the enzyme lactase, the lactose is not significantly hydrolyzed or absorbed. It remains in the intestine where it acts osmotically to draw water into the gut and serves as a substrate for conversion to lactic acid, CO_2, and H_2S by intestinal bacteria. The end result is bloating, flatulence, and diarrhea—all of which can be avoided simply by eliminating milk and milk products from the diet.

CLIN. CORR. 25.6 FAD DIETS: HIGH PROTEIN–HIGH FAT

One of the most popular fad diets over the years has been the high-protein/high-fat diet such as the Stillman and Atkins diets. Actually, at first glance, it sounds as if these diets should have a sound metabolic basis. The basic premise is that if one severely restricts carbohydrate intake, it is possible to eat large amounts of high-protein/high-fat foods because the body will not be able to efficiently utilize the fat. This hypothesis is primarily based on the fact, made abundantly clear in any biochemistry textbook, that glucose is needed to replenish the intermediates of the citric acid cycle. Thus, in the absence of glucose, fat should simply be converted to ketone bodies and be disposed of.

There are several problems with this oversimplified hypothesis. First, it ignores the fact that many amino acids can be readily converted to citric acid cycle intermediates. Second, many tissues in the body are perfectly able to use ketone bodies for energy generation. Third, the loss of ketones in the urine cannot possibly lead to any significant caloric deficit. Maximum ketone excretion is about 20 g (100 kcal)/day.

It is important, however, to realize that this diet does appear to "work" for many patients. The apparent success of the diet is related primarily to two factors. In the first place, any low carbohydrate diet results in a significant initial water loss. This is primarily due to depletion of glycogen reserves, since 3 g of water is bound for every 1 g of glycogen. It is this rapid initial weight loss which makes the diet so

appealing. Second, while a high-protein/high-fat diet sounds appealing initially, it is relatively unpalatable and expensive, leading ultimately to decreased caloric intake.

This diet is also not without its health risks. The high-fat content may contribute to atherosclerosis and heart disease. The lack of fruits and vegetables may lead to vitamin deficiencies. The ketone bodies could irritate the kidneys unless large amounts of water are consumed. Finally, ketosis should be avoided by pregnant women (ketone bodies can be harmful to the developing fetal brain), and high-protein intakes should be avoided by anyone with a history of liver or kidney disease.

CLIN. CORR. 25.7 CARBOHYDRATE LOADING AND ATHLETIC ENDURANCE

Much of the folklore concerning special nutritional requirements or specific nutritional regimes for athletes has little basis in fact. However, one common practice, that of carbohydrate loading, does appear to have biochemical backing. This technique is frequently used by track athletes to increase endurance. It consists of a 4- to 5-day period of heavy exercise while on a low carbohydrate diet, followed by three days of light exercise while on a high carbohydrate diet. The initial low carbohydrate-high energy demand period appears to cause a depletion of muscle glycogen stores. Apparently, the subsequent change to a high carbohydrate diet causes a slight rebound effect with the production of higher than normal levels of insulin and growth hormone. Under these conditions glycogen storage is favored and

25.7 FATS

Triglycerides, or fats, can be directly utilized by many tissues of the body as an energy source and are an important part of membrane structure. Any excess fat in the diet can be stored as triglyceride only. As with carbohydrate, the human body can adapt to a wide range of fat intakes. However, some problems can develop at the extremes (either high or low) of fat consumption. At the low end, essential fatty acid (EFA) deficiencies may become a problem. The fatty acids linoleic, linolenic, and arachidonic acid cannot be made by the body and thus are essential components of the diet. These EFA are needed for maintaining the function and integrity of membrane structure, for fat metabolism and transport, and for synthesis of prostaglandins. The most characteristic symptom of essential fatty acid deficiency is a scaly dermatitis. EFA deficiency is very rare in the United States, being seen primarily in low-birthweight infants fed on artificial formulas lacking EFA and in hospitalized patients maintained on total parenteral nutrition for long periods of time.

At the other end of the scale, there is ligitimate concern that excess fat in the diet does cause elevation of serum lipids and thus an increased risk of heart disease. Most experts agree with that general conclusion. Unfortunately, there is no firm consensus as to how much is too much. However, our understanding of metabolism tells us that any fat consumed in excess of energy needs (except for the small amount needed for membrane formation) has nowhere to go but our adipose tissue, and obesity is correlated with an increased risk of heart disease, diabetes, and stroke.

25.8 COMPOSITION OF MACRONUTRIENTS IN THE DIET

From the foregoing discussion it is apparent that there are relatively few instances of macronutrient deficiencies in the American diet. Thus, much of the interest in recent years has focused more on the question of whether there is an ideal diet composition consistent with good health. It would be easy to pass off such discussions as purely academic, yet our understanding of these issues could well be

vital. Heart disease, stroke, and cancer kill many Americans each year, and if some experts are even partially correct, many of these deaths could be preventable with prudent diet. So it is only fitting that we now turn to the question of diet composition.

Lipid Composition of the Diet

Most of the current discussion centers around two key issues: (1) Can serum cholesterol and triglyceride levels be controlled by diet? (2) Does lowering serum and triglyceride levels protect against heart disease? The controversies centered around dietary control of cholesterol levels illustrate perfectly the trap one falls into by trying to look too closely at each individual component of the diet instead of the diet as a whole. For example, there are at least four components that can be identified as having an effect on serum cholesterol: cholesterol itself, polyunsaturated fatty acids (PUFA), saturated fatty acids (SFA), and fiber. It would appear obvious that the more cholesterol one eats, the higher the serum cholesterol would be. However, cholesterol synthesis is tightly regulated via a feedback control at the hydroxymethylglutaryl CoA reductase step, so decreases in dietary cholesterol have relatively little effect on serum cholesterol levels (Chapter 10). One can obtain a more significant reduction in cholesterol and triglyceride levels by increasing the ratio of PUFA/SFA in the diet. While the effects of PUFA are more dramatic, the biochemistry of their action is still uncertain. Finally, some plant fibers, especially pectins from fruits and guar gums from legumes, appear to bind cholesterol and decrease its absorption.

Actually there is very little disagreement with respect to these data. The question is, what can be done with the information? Much of the disagreement arises from the tendency to look at each dietary factor in isolation. For example, it is indeed debatable whether it is worthwhile placing a patient on a highly restrictive 300-mg cholesterol diet (1 egg = 250 mg of cholesterol) if his serum cholesterol is lowered only 5–10%. Likewise, changing the PUFA/SFA ratio from 0.3 (the current value) to 1.0 would either require a radical change in the diet by elimination of foods containing saturated fat (largely meats and fats) or an addition of large amounts of rather unpalatable polyunsaturated fats to the diet. For many Americans this would be unrealistic. Fiber is another good example. One could expect, at the most, a 5% decrease in serum cholesterol by adding any reasonable amount of fiber to the diet. (Very few people would eat the 10 apples per day needed to lower serum cholesterol by 15%.) Are we to conclude that any dietary means of controlling cholesterol levels are useless then? Only if each element of the diet is examined in isolation. For example, recent studies have shown that vegetarians, who

glycogen stores can reach almost twice the normal amounts. This does significantly increase endurance, since exercising muscle relies primarily on glucose for its energy needs and glycogen is the most readily metabolizable form of glucose.

Test subjects on a high-fat and high-protein diet had less than 1.6 g of glycogen/100 g of muscle and could perform a standardized work load for only 60 min. When the same subjects then followed a high carbohydrate diet for 3 days, their glycogen stores increased to 4 g/100 g of muscle and the same workload could be performed for up to 4 h. With time, of course, the body would adapt to the high carbohydrate intake, with insulin and glycogen levels returning to normal. No potential side effects of this dietary regimen are known at present, although the diet will also cause deposition of water in the muscle. The combination of glycogen and water deposits could cause muscle stiffness and potential damage.

have lower cholesterol plus higher PUFA/SFA ratios and higher fiber intakes, may average 25–30% lower cholesterol levels than their nonvegetarian counterparts. Perhaps, more to the point, diet modification of the type acceptable to the average American has been shown to cause a 10–15% decrease in cholesterol levels in long-term studies.

The second question is much more difficult. Certainly, many epidemiologic studies have shown that populations with lower serum cholesterol levels run a lower risk of heart disease. However, this does not prove a cause and effect relationship. Animal studies have also shown that low cholesterol–high PUFA–high fiber diets decrease the incidence of coronary artery disease. However, these animal models might not be completely applicable to man. The more recent discovery that cholesterol in the form of HDL particles may actually play a protective role, complicates the picture further. Many of the dietary regimens which lower total serum cholesterol, lower HDL cholesterol as well as LDL cholesterol. Finally, it is important to keep in mind that diet is just one of many risk factors. Unfortunately, some nutritionists, and certainly the popular press, have overemphasized the importance of this one risk factor in preventing heart disease.

Carbohydrates

Much of the current dispute in the area of carbohydrates centers around the amount of refined carbohydrate in the diet. It is possible in the popular press to see simple sugars (primarily sucrose) blamed for almost every ill from tooth decay to heart disease and diabetes. In the case of tooth decay, these assertions are clearly correct. In the case of heart disease and diabetes, however, the linkage is more obscure.

It is evident that much of the excess dietary carbohydrate in the American diet is converted to triglyceride in the liver, exported as VLDL, and stored in the adipose tissue. It is even somewhat logical to assume that simple sugars, which are absorbed and metabolized very rapidly, might cause a slightly greater elevation of triglycerides than complex carbohydrates. In actual studies with human volunteers, an isocaloric switch from a diet high in starch to one high in simple sugars, does cause a transient rise in triglyceride levels. However, over a period of 2–3 months, adaptation occurs and the triglyceride levels return to normal. Thus, for most individuals, there is no evidence that simple sugars can cause a permanent elevation of serum triglycerides. However, among individuals with hypertriglyceridemia a certain fraction are very sensitive to carbohydrate intake. Individuals with this carbohydrate-induced hypertrigly-

ceridemia do respond well to dietary restriction of simple sugars, and thus it is often worthwhile to attempt dietary modification for individuals with elevated triglycerides.

The situation with respect to diabetes is probably even less direct. Whereas restriction of simple sugars is almost always desirable in a patient who already has diabetes, there is little direct evidence that an excess of simple sugars in the diet is a direct cause of diabetes. However, foods rich in simple sugars do have a very high caloric density, contributing to overeating and obesity. Obesity, as discussed earlier, does have a direct relationship to heart disease and diabetes. This may go a long way toward explaining some of the epidemiologic studies linking consumption of simple sugars with heart disease and diabetes.

Protein

Much concern has also been voiced recently about the type of protein in the American diet. Epidemiologic data and animal studies suggest that consumption of animal protein is associated with increased incidence of heart disease and various forms of cancer. One would assume that it is probably not the animal protein itself that is involved, but the associated fat and cholesterol. What sort of protein should we consume? Although the present diet may not be optimal, it would be very difficult, if not impossible, to prove that animal protein consumption causes heart disease or cancer. Furthermore, a strictly vegetarian diet is not without some health risks of its own, unless the individual is nutritionally very well informed (Clin. Corr. 25.1). Perhaps a middle road is best. Clearly there are no known health dangers associated with a mixed diet that is lower in animal protein than the current American standard.

Recommendations

In the midst of all of this controversy, it would seem to be premature to make specific recommendations with respect to the ideal dietary composition for the American public. Yet that is just what several private and government groups have done in recent years. This movement was spearheaded by the Senate Select Committee on Human Nutrition which first published its *Dietary Goals for the United States* in 1977. The Senate Select Committee recommended that the American public reduce consumption of total calories, total fat, saturated fat, cholesterol, simple sugars, and salt to "ideal" goals more compatible with good health (Figure 25.3). Subsequently, the USDA and HEW have published recommendations that are similar, but are set in the format of general guidelines. In effect,

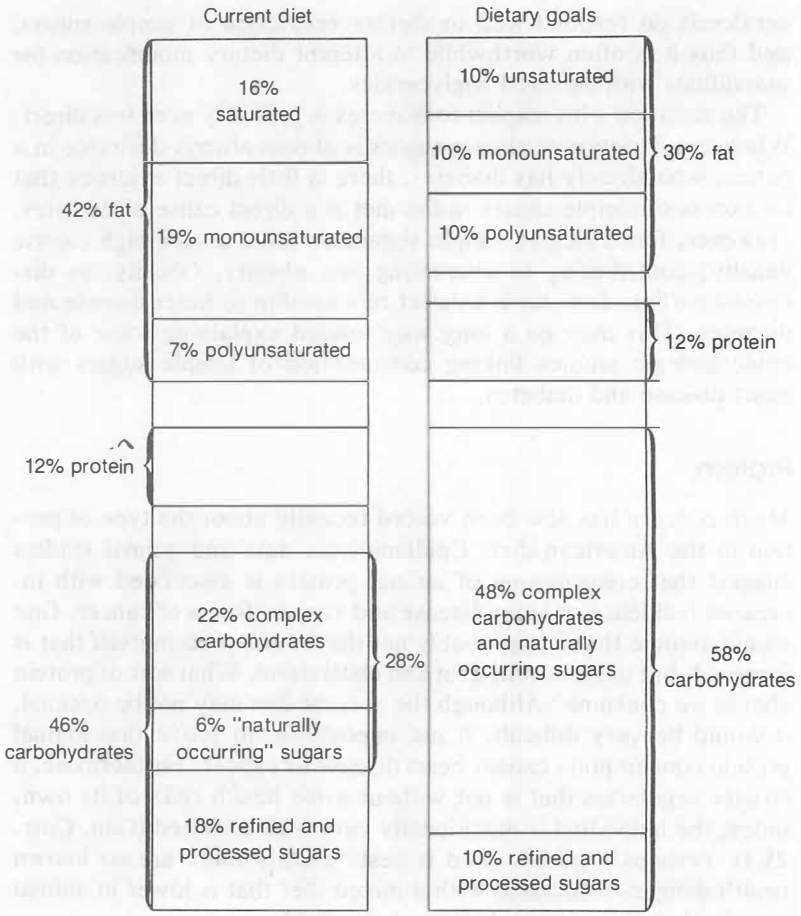


Figure 25.3
U.S. dietary goals.
Graphical comparison of the composition of the current U.S. diet and the dietary goals for the U.S. population suggested by the Senate Select Committee on Human Nutrition.
From *Dietary Goals for the United States, 2nd ed.*. U.S. Government Printing Office, Washington, D.C., 1977.

these agencies agreed that the dietary guidelines recommended by the Senate Select Committee were beneficial, but rejected the idea of rigid goals because of individual variability. The Food and Nutrition Board of the National Research Council, on the other hand, has published guidelines that differ strongly with respect to recommendations on fat and cholesterol intake. They concluded that the evi-

dence was not yet convincing that reduction of fat and cholesterol were beneficial. It is clear from the controversy generated by these various guidelines that it may be too soon to generate firm dietary goals for everyone in the population.

The most important argument against such recommendations is that we presently do not have enough information to set concrete goals. We might, for example, be creating some problems while solving others. For example, the goals of reducing total fat and saturated fat in the diet are best met by replacing animal protein with vegetable protein. This, in turn, might reduce the amount of available iron and vitamin B_{12} in the diet. It is also quite clear that the same set of guidelines do not apply for every individual. For example, exercise is known to raise HDL cholesterol and obesity is known to elevate cholesterol, triglycerides, and reduce glucose tolerance. Thus the very active individual who maintains ideal body weight can likely tolerate higher fat and sugar intakes than an obese individual.

On the "pro" side, however, it clearly can be argued that all of the dietary recommendations are in the right direction for reducing nutritional risk factors in the general population. Furthermore, similar diets have been consumed by our ancestors and by people in other countries with no apparent harm. Whatever the outcome of this debate in the years ahead, it will undoubtedly shape much of our ideas concerning the role of nutrition in medicine.

BIBLIOGRAPHY

Starvation

Cahill, G. F., Jr. Starvation in man. *N. Engl. J. Med.* 282:668, 1970.

Young, V. R., and Scrinshaw, N. S. The physiology of starvation. *Sci. Am.* 225:214, 1971.

Occurrence and Management of Protein Energy Malnutrition

Bistrian, B. R. Interaction of nutrition and infection in the hospital setting. *Am. J. Clin. Nutr.* 30:1228, 1977.

Blackburn, G. L., and Bistrian, B. R. Curative nutrition: protein-caloric management, in H. Schneider, C. Anderson, and D. Coursin (eds.), *Nutritional support of medical practice*. New York: Harper & Row, 1977, p. 80.

Flatt, J. P., and Blackburn, G. L. The metabolic fuel regulatory system: implications for protein sparing therapies during caloric deprivation and disease. *Am. J. Clin. Nutr.* 27:175, 1974.

Metabolic Consequences of Overweight

Gordon, E. S. Metabolic aspects of obesity. *Adv. Metab. Disord.* 4:229, 1970.

Sims, E. A., Horton, E. S., and Salans, L. B. Inducible metabolic abnormalities during development of obesity. *Annu. Rev. Med.* 22:235, 1971.

Vegetarian Diets

Lappe, F. M. *Diet for a small planet.* New York: Ballantine, 1975.

Raper, N. R., and Hill, M. M. Vegetarian diets. *Nutr. Rev.* 32:29, 1974.

Fad Weight Loss Diets

Berland, T. *Rating the diets.* New York: Signet, 1979.

Macronutrient Composition and Health

Glueck, C. J., and Conner, W. E. Diet-coronary heart disease relationships. *Am. J. Clin. Nutr.* 31:727, 1978.

Reiser, R. Oversimplification of diet-coronary heart disease relationships. *Am. J. Clin. Nutr.* 31:865, 1978.

Symposium on dietary carbohydrate in man. *Am. J. Clin. Nutr.* 20:65, 1967.

Symposium on nutrition in the causation of cancer. *Cancer Res.* 35:3235, 1975.

Dietary Recommendations

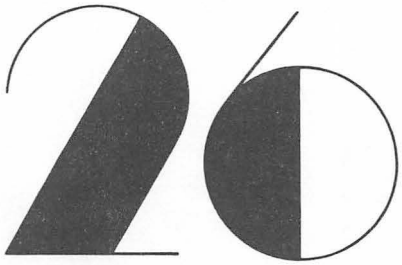
Harper, A. E. Arguments against U.S. dietary goals. *J. Nutr. Ed.* 9:154, 1977.

Hegsted, D. M. Dietary goals—a progressive view. *Am. J. Clin. Nutr.* 31:1504, 1978.

Food and Nutrition Board of the National Academy of Sciences. *Towards healthful diets.* Washington, D.C.: U.S. Government Printing Office, 1980.

Senate Select Committee on Human Nutrition. *Dietary goals for the United States, 2nd ed.*, Stock no. 052-070-04376-8. Washington, D.C.: U.S. Government Printing Office, 1977.

U.S. Department of Agriculture. *Nutrition and your health, dietary guidelines for Americans.* Stock no. 017-001-00416-2. Washington, D.C.: U.S. Government Printing Office, 1980.



Principles of Nutrition II

STEPHEN G. CHANEY

MICRONUTRIENTS

The micronutrients play a vital role in human metabolism, since they are involved in almost every biochemical reaction and pathway known to man. However, the biochemistry of these nutrients is of little interest unless we also know if dietary deficiencies are likely. Alarming reports of nutritional deficiencies in the American diet continually appear. Is there any truth to these predictions of doom? On the one hand, the American diet is undoubtedly the best that it ever has been. Our current food supply provides us with an abundant variety of foods all year long—a luxury not available in the "good old days"—and deficiency diseases have become medical curiosities. On the other hand, our diet is far from optimal. The old adage that we get everything we need from a balanced diet is true only if in fact we eat a balanced diet. Unfortunately, most Americans do not know how to select a balanced diet. Foods of high caloric density and low nutrient density (often referred to elsewhere as empty calories or junk food) are an abundant and popular part of the American diet. Obviously then, neither alarm nor complacency are fully justified. We need to know how to best evaluate the adequacy of our diet.

26.1 ASSESSMENT OF MALNUTRITION

Why does one see so many reports of vitamin and mineral deficiencies when deficiency diseases are so rare? In many cases these reports result from the misinterpretation of valid scientific data. One needs to be aware that there are three increasingly stringent criteria for measuring malnutrition. (1) Dietary intake studies, which are usually based on a 24-hour recall, are the least stringent. In the first place, 24-hour recalls almost always tend to overestimate the number of people with deficient diets. Also, poor dietary intake alone is usually not a problem in this country unless the situation is compounded by increased need. Thus dietary surveys are not indicative of malnutrition by themselves, although they are often quoted as if they were. (2) Biochemical assays, either direct or indirect, are a more useful indicator of the nutritional status of an individual. At their best, they can indicate subclinical nutritional deficiencies, which can be treated before actual deficiency diseases develop. However, all biochemical assays are not equally valid—an unfortunate fact that is not sufficiently recognized. Furthermore, changes in biochemical parameters due to stress need to be interpreted with caution. The distribution of many nutrients in the body changes

dramatically in a stress situation such as illness, injury, and pregnancy. A drop of that nutrient in one tissue compartment (usually blood) need not signal a deficiency or an increased requirement. It could simply reflect a normal metabolic adjustment to stress. (3) The most stringent criterion is, of course, the appearance of clinical symptoms. Obviously, any time clinical symptoms of a deficiency are present, especially if they respond to diet, it is a serious concern.

The question remains: When should dietary surveys or biochemical assays be interpreted to indicate the necessity of nutritional intervention? At what level should we become concerned? Obviously, the situation is complex and controversial, but the following general guidelines are probably useful: Dietary surveys are seldom a valid indication of general malnutrition unless the average intake for a population group falls significantly below the standard (usually two-thirds of the Recommended Dietary Allowance) for that nutrient. This seldom occurs in the American population. However, by looking at the number of individuals with intakes below the standard, it is possible to identify high-risk population groups that should be monitored more closely. This is the real value of dietary surveys. Biochemical assays can indicate subclinical cases of malnutrition where nutritional intervention is desirable provided (a) the assay has been shown to be reliable, (b) the deficiency can be verified by a second assay, and (c) there is no unusual stress s tuation that may alter micronutrient distribution. In evaluating nutritional claims and counterclaims it is well to keep an open, but skeptical, mind and evaluate each issue on the basis of its scientific merit. While serious vitamin and mineral deficiencies are rare in this country, mild to moderate deficiencies can be found in certain select population groups. It is important for the clinician to be aware of these population groups at risk and their most probable symptoms.

26.2 RECOMMENDED DIETARY ALLOWANCES

One hears a lot about the Recommended Dietary Allowances (RDA). What are they and how are they determined? Briefly, "The Recommended Dietary Allowances are the levels of intake of essential nutrients considered in the judgement of the Food and Nutrition Board of the National Research Council on the basis of available scientific knowledge, to be adequate to meet the known nutritional need of practically all healthy persons." Optimally, the RDAs are based on the minimum daily intake sufficient to prevent the appear-

ance of nutritional deficiency in 95% of the population. This determination is relatively easy to make for those nutrients associated with dramatic deficiency diseases, such as vitamin C and scurvy. In other instances more indirect measures must be used, such as tissue saturation or extrapolation from animal studies. In some cases, such as vitamin E, in which no deficiency symptoms are known to occur in the general population, the RDA is simply defined as the normal level of intake in the American diet. Obviously, there is no one set of criteria that can be used for all micronutrients, and there is always some uncertainty and debate as to the correct criteria. Furthermore, the criteria are constantly changed by new research. The Food and Nutrition Board meets every 5 years (most recently in 1979) to consider currently available information and update their recommendations.

The RDAs serve as a useful general guide in evaluating the adequacy of diets and (as the USRDA) the nutritional value of foods. However, the RDAs have several limitations that should be kept in mind. Some of the most important limitations are as follows: (1) The RDAs represent an ideal average intake for groups of people and are best used for evaluating nutritional status of population groups. The RDAs are not meant to be standards or requirements for individuals. Some individuals would have no problem with intakes below the RDA, whereas others might develop deficiencies on intakes above the RDA. (2) The RDAs were designed to meet the needs of healthy people and do not take into account any special needs arising from infections, metabolic disorders, or chronic diseases. (3) Since present knowledge of nutritional needs is incomplete, there may be unrecognized nutritional needs. To provide for these needs, the RDAs should be met from as varied a selection of foods as possible. No single food can be considered complete, even if it meets the RDA for all known nutrients. This is an important consideration, especially in light of the current practice of fortifying foods of otherwise low nutritional value. (4) The RDAs make no effort to define the "optimal" level of any nutrient, since optimal levels are almost impossible to define on the basis of current scientific information.

26.3 FAT-SOLUBLE VITAMINS

Vitamin A

The active forms of vitamin A are retinol, retinaldehyde, and retinoic acid. These substances are synthesized by plants as the

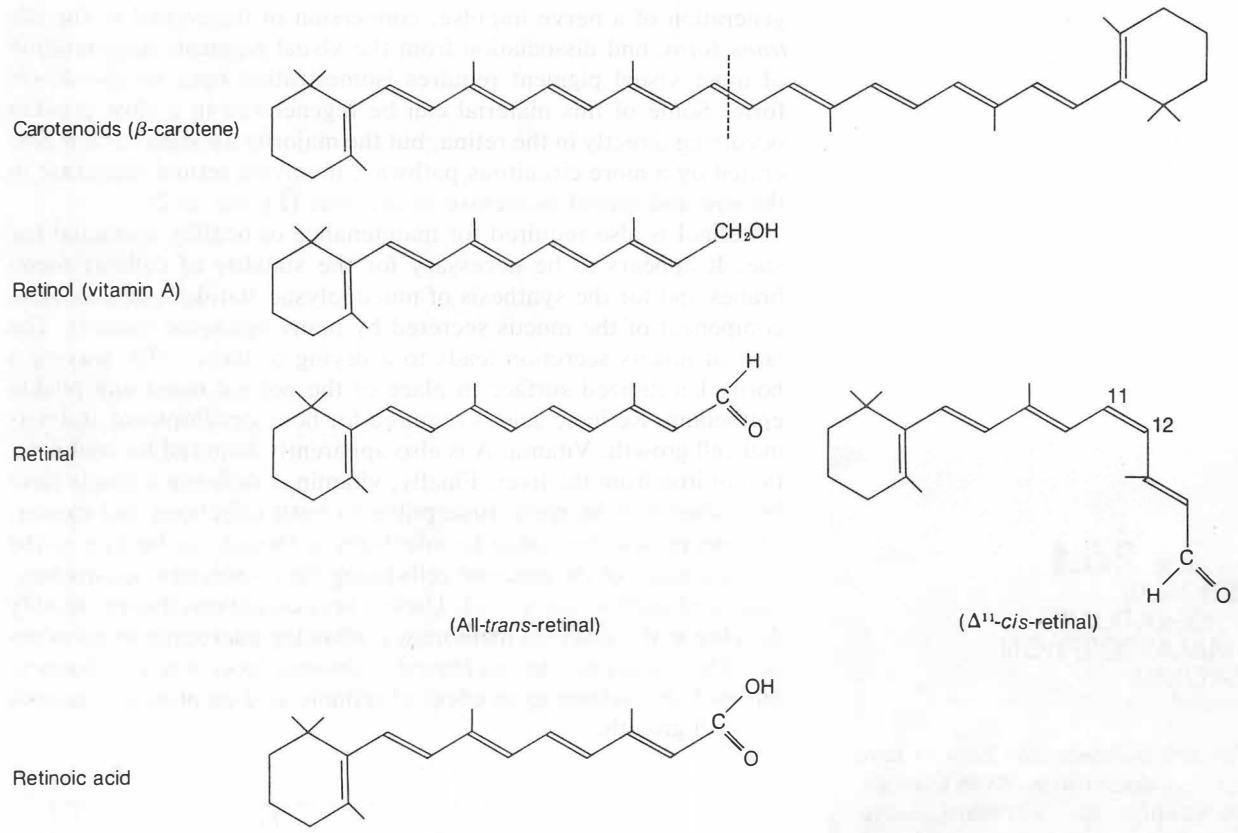


Figure 26.1
Structures of vitamin A and related compounds.

more complex carotenoids (Figure 26.1), which are cleaved to retinol by most animals and stored in the liver as retinol palmitate. Liver, egg yolk, butter, and whole milk are good sources of the preformed retinol. Dark green and yellow vegetables are generally good sources of the carotenoids. The conversion of carotenoids to retinol is rarely 100%, so that the vitamin A potency of various foods is expressed in terms of retinol equivalents (1 RE is equal to 1 μg retinol, 6 μg β-carotene, and 12 μg of other carotenoids).

Vitamin A plays a number of important roles in the body. However, only in the case of vision is the biochemistry well understood. Vitamin A becomes reversibly associated with the visual pigments in the Δ^{11}-*cis*-retinal form. When light strikes the retina, a number of complex biochemical changes take place, resulting in the

generation of a nerve impulse, conversion of the retinal to the all-*trans* form, and dissociation from the visual pigment. Regeneration of more visual pigment requires isomerization back to the Δ^{11}-*cis* form. Some of this material can be regenerated in a slow process occurring directly in the retina, but the majority appears to be regenerated by a more circuitous pathway, involving retinal reductase in the eye and retinal isomerase in the liver (Figure 26.2).

Retinol is also required for maintenance of healthy epithelial tissue. It appears to be necessary for the stability of cellular membranes and for the synthesis of mucopolysaccharides (an important component of the mucus secreted by many epithelial tissues). The lack of mucus secretion leads to a drying of these cells, leaving a horny keratinized surface in place of the normal moist and pliable epithelium. Retinoic acid is required for bone development and normal cell growth. Vitamin A is also apparently required for mobilization of iron from the liver. Finally, vitamin A-deficient animals have been shown to be more susceptible to both infections and cancer. The decreased resistance to infections is thought to be due to the keratinization of the mucosal cells lining the respiratory, gastrointestinal, and genitourinary tract. Under these conditions fissures readily develop in the mucosal membranes, allowing microorganisms to enter. The reason for the decreased resistance to cancer is unknown, but probably relates to an effect of retinoic acid on protein synthesis and cell growth.

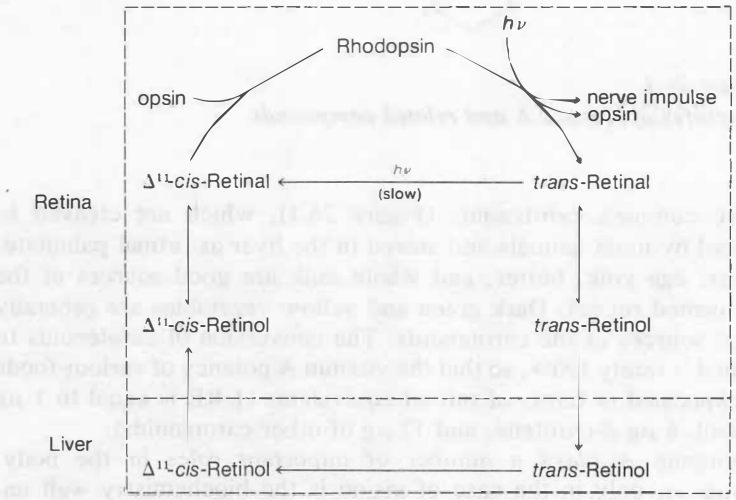


Figure 26.2
The role of vitamin A in vision.

CLIN. CORR. 26.1 NUTRITIONAL CONSIDERATIONS FOR MALABSORPTION SYNDROMES

Patients with malabsorption diseases have a high risk of malnutrition. As an example, let us examine the nutritional consequences of two diseases with malabsorption components. Cystic fibrosis (CF) involves a generalized dysfunction of the exocrine glands that leads to formation of a viscid mucus, which progressively plugs the ducts. Obstruction of the bronchi and bronchioles leads to pulmonary infections, which are usually the direct cause of death. However, in many cases the exocrine glands of the pancreas are also affected, leading to a deficiency of pancreatic enzymes and often a partial obstruction of the common bile duct.

The deficiency (or partial deficiency) of pancreatic lipase and bile salts leads to severe malabsorption of fat and fat-soluble vitamins. Calcium tends to form insoluble salts with the long-chain fatty acids, which accumulate in the intestine. While these

Since vitamin A is stored by the liver, deficiencies of this vitamin can develop only over prolonged periods of inadequate uptake. Mild vitamin A deficiencies are characterized by follicular hyperkeratosis (rough keratinized skin resembling "goosebumps"), anemia (biochemically equivalent to iron deficiency anemia, but in the presence of adequate iron intake), and increased susceptibility to infection. Night blindness is also an early symptom of vitamin A deficiency. Severe vitamin A deficiency leads to a progressive keratinization of the cornea of the eye. This condition is known as Bitot's spots in its mildest form (when the keratinized areas are still localized as distinct spots), xerosis conjunctivae in moderately severe form, and as xerophthalmia in its most advanced stages. In the final stages, infection usually sets in, with resulting hemorrhaging of the eye and permanent loss of vision.

The severe symptoms of vitamin A deficiency are generally seen only in developing countries. In this country even mild vitamin A deficiencies are rare, but the potential for deficiencies does exist. For most people (unless they happen to eat liver) the dark green and yellow vegetables are the most important dietary source of vitamin A. Unfortunately, these are the foods most often missing from the American diet. Nationwide, dietary surveys indicate that between 40 and 60% of the population consumes less than two-thirds of the RDA for vitamin A. The important question is how significant these dietary surveys are. While plasma vitamin A levels are low in a significant number of individuals, the clinical symptoms of vitamin A deficiency are rare. Follicular hyperkeratosis is occasionally seen and is the most characteristic symptom of vitamin A deficiency. The anemia and decreased resistance to infection also occur early, but they are too nonspecific to be useful indicators of vitamin A status. Night blindness is seldom seen in the general population. While clinically detectable vitamin A deficiency is rare in the general population, it is a fairly common consequence of severe liver damage or diseases that cause fat malabsorption (Clin. Corr. 26.1).

Recent studies have suggested that certain retinoids may be therapeutically useful in the treatment of epithelial cancers. However, this does not mean that vitamin A itself will ever be useful in the treatment of cancer. Vitamin A is only slowly converted to retinoic acid by the body, and unlike retinoic acid, can accumulate to toxic levels in the liver.

Since vitamin A does accumulate in the liver, large amounts of this vitamin over prolonged periods of time can be toxic. Doses of 15,000 to 50,000 RE per day over a period of months or years will prove to be toxic for many children and adults. The usual symptoms include bone pain, scaly dermatitis, enlargement of liver and spleen, nausea, and diarrhea. It is, of course, virtually impossible to ingest

are the most severe problems, some starches and proteins are also trapped in the fatty bolus of partially digested foods. This physical entrapment, along with the deficiencies of pancreatic amylase and pancreatic proteases, can lead to protein-calorie malnutrition as well. In the treatment of these patients, the total fat content of the diet is decreased drastically and medium-chain triglycerides (MCTs) used as a partial replacement, since they can be absorbed directly through the intestinal mucosa in the absence of bile salts and pancreatic lipase. Total calories in a CF diet are increased to compensate for the inefficiency of digestion, with most of the calories coming from simple sugars. Protein nutrition is improved with a high protein diet [3–5 g/(kg · day)] and pancreatic extracts taken orally with meals. Water-soluble forms of vitamins A, D, E, and K are usually included in the diet as well.

Gluten-sensitive enteropathy (celiac disease or nontropical sprue) appears to be caused by an immunologic reaction to the protein gluten found in wheat, rye, oats, and barley. This immunologic reaction causes atrophy of the intestinal villi, resulting in loss of surface area, brush border enzymes, and protein carriers. Secondary to the villous atrophy is impaired production of the hormone cholecystokinin-pancreozymin, which normally stimulates pancreatic enzyme and bile salt secretion. Thus many of the nutritional problems are the same as with the CF patient. However, the loss of brush border enzymes and protein carriers leads to more severe protein and calcium malabsorption, and may cause malabsorption of folic acid and iron. A gluten-free diet is, of course, the most important therapy for such a patient. Nutritional problems, when they do arise, can be handled in much the same manner as for the CF patient. Calcium, iron, and folic acid supplements may also be important.

toxic amounts of vitamin A from normal foods unless one eats polar bear liver (6,000 RE/g) regularly. Most instances of vitamin A toxicity are due to the use of massive doses of this vitamin to treat acne or prevent colds. Fortunately, this practice, while once common, is now relatively rare.

Vitamin D

Technically, vitamin D is a hormone rather than a vitamin. Cholecalciferol (D_3) is produced in the skin by ultraviolet irradiation of 7-dehydrocholesterol, a normal metabolite of cholesterol (see Figure 26.3). Thus, as long as the body is exposed to adequate sunlight,

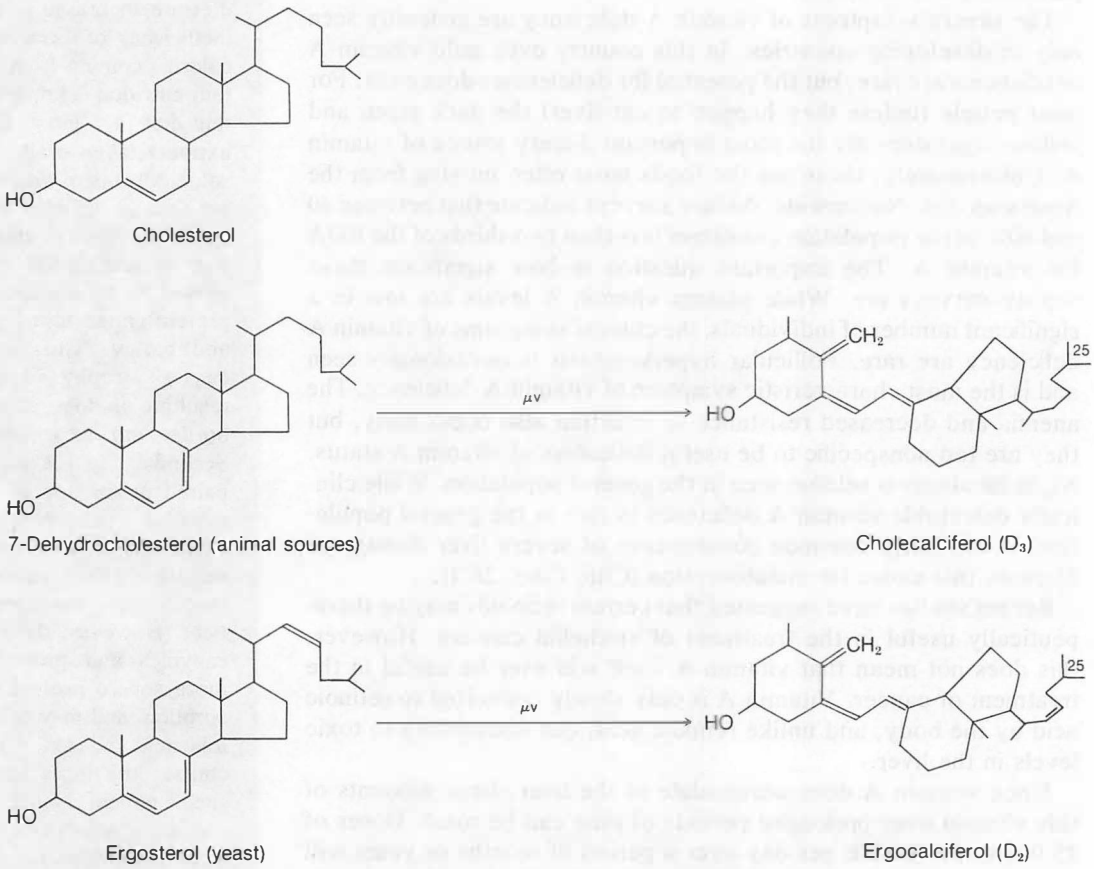


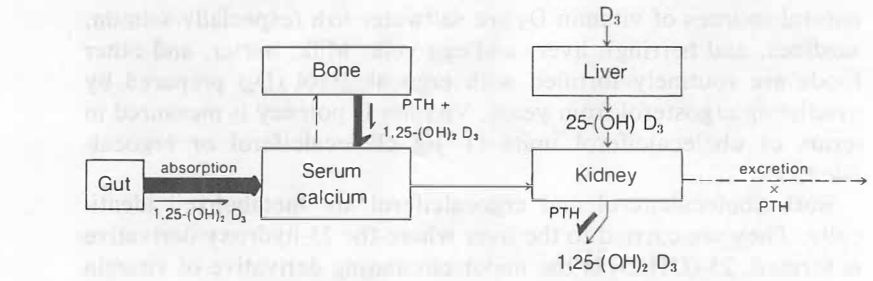
Figure 26.3
Structures of vitamin D and related compounds.

there is little or no dietary requirement for vitamin D. The best natural sources of vitamin D_3 are saltwater fish (especially salmon, sardines, and herring), liver, and egg yolk. Milk, butter, and other foods are routinely fortified with ergocalciferol (D_2) prepared by irradiating ergosterol from yeast. Vitamin D potency is measured in terms of cholecalciferol units (1 μg cholecalciferol or ergocalciferol).

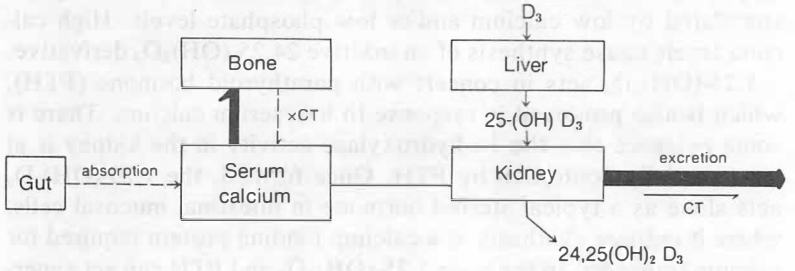
Both cholecalciferol and ergocalciferol are metabolized identically. They are carried to the liver where the 25-hydroxy derivative is formed. 25-(OH)D_3 is the major circulating derivative of vitamin D_3, and it is in turn converted into the biologically active 1α,25-dihydroxy derivative in the kidney. The synthesis of 1,25-$(OH)_2D_3$ is stimulated by low calcium and/or low phosphate levels. High calcium levels cause synthesis of an inactive 24,25-$(OH)_2D_3$ derivative.

1,25-$(OH)_2D_3$ acts in concert with parathyroid hormone (PTH), which is also produced in response to low serum calcium. There is some evidence that the 1α-hydroxylase activity in the kidney is at least partially controlled by PTH. Once formed, the 1,25-$(OH)_2D_3$ acts alone as a typical steroid hormone in intestinal mucosal cells, where it induces synthesis of a calcium binding protein required for calcium transport. In the bone 1,25-$(OH)_2D_3$ and PTH can act synergistically to promote resorption (demineralization) by stimulating osteoblast formation and activity. Finally, PTH inhibits calcium excretion in the kidney. The overall response of calcium metabolism to several different physiological situations is summarized in Figure 26.4. The response to low serum calcium levels is characterized by elevation of PTH and 1,25-$(OH)_2D_3$, which act to enhance calcium absorption and bone resorption and inhibit calcium excretion (Figure 26.4*a*). High serum calcium levels cause production of the hormone calcitonin, which inhibits bone resorption and enhances calcium excretion. Furthermore, the high levels of serum calcium and phosphate increase the rate of bone mineralization (Figure 26.4*b*). Thus bone serves as a very important reservoir of the calcium and phosphate needed to maintain homeostasis of serum levels. When vitamin D and dietary calcium are adequate, no net loss of bone calcium occurs. However, when dietary calcium is low, PTH and 1,25-$(OH)_2D_3$ will cause net demineralization of bone to maintain normal serum calcium levels. Vitamin D deficiency also causes net demineralization of bone due to elevation of PTH (Figure 26.4*c*).

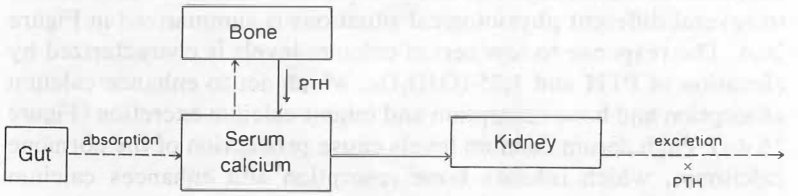
The most common symptoms of vitamin D deficiency are rickets in young children and osteomalacia in adults. Rickets is characterized by continued formation of osteoid matrix and cartilage, which are improperly mineralized resulting in soft, pliable bones. In the adult demineralization of preexisting bone takes place, causing the bone to become softer and more susceptible to fracture. This


D3
Bone
Liver
PTH +
1,25-(OH)2 D3
25-(OH) D3
Gut
absorption
Serum calcium
Kidney
excretion
×
PTH
1,25-(OH)2 D3
PTH
1,25-(OH)2 D3

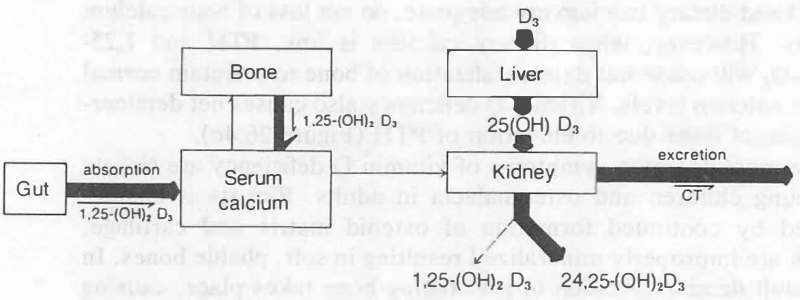
(*a*)


D3
Bone
Liver
×CT
25-(OH) D3
Gut
absorption
Serum calcium
Kidney
excretion
CT
24,25(OH)2 D3

(*b*)


Bone
PTH
Gut
absorption
Serum calcium
Kidney
excretion
×
PTH

(*c*)


D3
Bone
Liver
1,25-(OH)2 D3
25(OH) D3
Gut
absorption
1,25-(OH)2 D3
Serum calcium
Kidney
excretion
CT
1,25-(OH)2 D3
24,25-(OH)2D3

(*d*)

osteomalacia is easily distinguishable from the more common osteoporosis, by the fact that the osteoid matrix remains intact in the former, but not in the latter.

Because of fortification of dairy products with vitamin D, dietary deficiencies are very rare. The cases of dietary vitamin D deficiency that do occur are most often seen in low-income groups with poor dietary habits, the elderly (who often also have minimal exposure to sunlight), strict vegetarians (especially if their diet is also low in calcium and high in fiber), and chronic alcoholics. Most cases of vitamin D deficiency, however, are a result of diseases causing fat malabsorption or severe liver and kidney disease (Clin. Corr. 26.2). Certain drugs also interfere with vitamin D metabolism. Anticonvulsant drugs inhibit the 25-hydroxylation reaction in the liver. Corticosteroids also have a similar effect and have been shown to cause bone demineralization when used as part of long-term therapy.

Vitamin D can also be toxic in doses 10–100 times the RDA. The mechanism of vitamin D toxicity is summarized in Figure 26.4*d*. Enhanced calcium absorption and bone resorption cause hypercalcemia, which can lead to metastatic calcifications. The enhanced bone resorption also causes bone demineralization similar to that seen in vitamin D deficiency. Finally, the high serum calcium leads directly to hypercalciuria which predisposes the patient to formation of renal stones.

Vitamin E

For many years vitamin E was described as the "vitamin in search of a disease." While vitamin E deficiency diseases are still virtually unknown, its metabolic role in the body has become better understood in recent years. Vitamin E as it occurs in the diet is a mixture of several closely related compounds, called tocopherols. α-Tocopherol is considered the most potent of these and is used as

Figure 26.4
Vitamin D and calcium homeostasis.
(a) Low serum calcium. (b) High serum calcium. (c) Low vitamin D. (d) Excess vitamin D. The dominant pathways of calcium metabolism under each set of metabolic conditions are shown with heavy arrows. The effect of various hormones on these pathways is shown by $\longrightarrow$ *(stimulation) or* $\times$ *(repression). PTH = parathyroid hormone, CT = calcitonin, D_3 = cholecalciferol, 25-$(OH)D_3$ = 24-hydroxycholecalciferol, 1,25-$(OH)_2D_3$ = 1α,25-dihydroxycholecalciferol.*

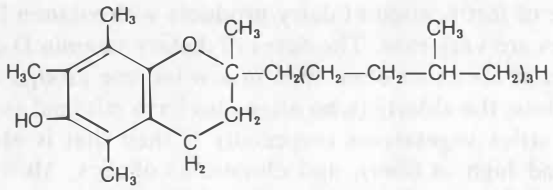
CLIN. CORR. 26.2 VITAMIN D-RESISTANT RICKETS

Classical rickets and osteomalacia can be cured by vitamin D supplementation. However, occasional cases of rickets arise which are resistant to vitamin D. Vitamin D-resistant rickets can be due to any number of unrelated causes. The most common form is hypophosphatemic vitamin D refractory rickets which results from general renal tubular dysfunction or from an inability to convert D_3 to 25-$(OH)D_3$ or 1,25-$(OH)_2D_3$. The symptoms of this disease are retarded growth and bone deformities similar to that seen in classical rickets. The traditional treatment has been oral administration of massive doses (50,000–500,000 IU) of vitamin D_2, but even these levels of vitamin D are seldom completely effective in treating the disease and some growth retardation is inevitable. More recently, synthetically prepared vitamin D metabolites are being used. Obviously 1,25-$(OH)_2D_3$ is the metabolite of choice, but it is currently available in only limited quantities. Thus, other metabolites, usually 25-$(OH)D_3$ or 1-$(OH)D_3$, are more commonly used. Both are reasonably effective, but when these metabolites are used, the normal feedback controls are not fully operative and serum calcium levels must be carefully monitored to prevent hypercalcemia from developing.

In chronic renal failure, a somewhat more complicated chain of events is set in motion, leading to a condition known as renal osteodystrophy. The renal faiiure results in an inability to produce the 1,25-$(OH)_2D_3$, and thus bone calcium becomes the only important source of serum calcium. The situation is complicated further by increased renal retention of phosphate

and a resulting hyperphosphatemia. The serum phosphate levels are often high enough to cause metastatic calcifications, which tends to lower serum calcium levels further (the solubility product of calcium phosphate in the serum is very low and a high serum level of one component necessarily causes a decreased concentration of the other). The hyperphosphatemia and hypocalcemia stimulate parathyroid hormone secretion, and the resulting hyperparathyroidism further accelerates the rate of bone loss. One ends up with both bone loss and metastatic calcifications. In this case, simple administration of high doses of vitamin D or its active metabolites would not be sufficient, since the combination of hyperphosphatemia and hypercalcemia would only lead to more extensive metastatic calcification. The readjustment of serum calcium levels by high calcium diets and/or vitamin D supplementation must be accompanied by phosphate reduction therapies. The most common technique is to use phosphate binding resins.

the measure of vitamin E potency (1α-tocopherol equiv = 1 mg α-tocopherol).



α-Tocopherol (5,7,8-trimethyltocol)

First and foremost, vitamin E appears to play an important role as a naturally occurring antioxidant. Due to its lipophilic structure it tends to accumulate in circulating lipoproteins, cellular membranes, and fat deposits, where it reacts very readily with molecular oxygen and free radicals. It acts as a scavenger for these compounds, protecting unsaturated fatty acids (especially those in the membranes) from peroxidation reactions. Vitamin E appears to play a role in cellular respiration, either by stabilizing coenzyme Q or by helping transfer electrons to coenzyme Q. It also appears to enhance heme synthesis by increasing the levels of δ-aminolevulinic acid (ALA) synthetase and ALA dehydratase. Most of these vitamin E effects are thought to be an indirect effect of its antioxidant potential, rather than its actual participation as a coenzyme in any biochemical reactions.

Symptoms of vitamin E deficiency vary widely from one animal species to another. In various animals vitamin E deficiencies can be associated with sterility, muscular dystrophy, central nervous system changes, and megaloblastic anemia. In humans, however, the symptoms are limited to increased fragility of the red blood cell membrane (presumably due to peroxidation of membrane components). Premature infants fed on formulas low in vitamin E often develop a form of hemolytic anemia that can be corrected by vitamin E supplementation. Adults suffering from fat malabsorption show a decreased red blood cell survival time, but seldom develop anemia itself. Hence, vitamin E supplementation is often necessary with premature infants and in cases of fat malabsorption.

Studies on the recommended levels of vitamin E in the diet have been hampered by the difficulty in producing severe vitamin E deficiencies in man. In general it is assumed that the vitamin E levels in the American diet are sufficient, since no major vitamin E deficiency diseases have been found. However, vitamin E requirements do increase as the intake of polyunsaturated fatty acids (PUFA) increases. While the recent emphasis on high PUFA diets to reduce serum cholesterol may be of benefit in controlling heart disease, the

propensity of PUFA to form free radicals on exposure to oxygen may lead to an increased cancer risk. Thus it appears only prudent to increase vitamin E intake along with high PUFA diets.

Although vitamin E deficiencies are rare, use of vitamin E in therapeutic doses is fairly widespread. In the past vitamin E has been promoted as a cure for so many diseases that it approached the status of a patent medicine. Many of the most heavily promoted uses of vitamin E were completely unfounded. Unfortunately, the resulting controversy has obscured more recent discoveries about vitamin E. For example, animal studies have suggested that 200 mg of vitamin E/day may protect lung tissue from the damaging effects of ozone in high smog areas. A regimen of 300–400 mg of vitamin E/day for a period of at least 3–6 months has been shown to alleviate a condition known as intermittent claudication (leg pain when walking due to poor circulation). Similar levels of vitamin E inhibit one of the oxidation steps leading to synthesis of certain prostaglandins. This appears to cause mild antiinflammatory properties and a prolongation of clotting time. The latter property has recently raised some interest, since vitamin E inhibits the same cyclooxygenase reaction as aspirin. However, no studies have been carried out to show whether vitamin E can be effective in reducing the recurrence of myocardial infarctions. This brings us to perhaps the most controversial use of vitamin E—to prevent and/or treat heart disease. None of the double-blind studies carried out to date have shown any significant effect of vitamin E. However, most of these studies have been criticized on the basis of the relatively low doses and short time periods involved.

As a fat-soluble vitamin, E has the potential for toxicity. However, it does appear to be the least toxic of the fat-soluble vitamins. No instances of toxicity have been reported at doses of 300 mg/day or less. A few scattered reports of malaise and easy fatiguability have been reported at doses of 800 mg/day.

Vitamin K

Vitamin K is found naturally as K_1 (phytylmenaquinone) in green vegetables and K_2 (multiprenylmenaquinone), which is synthesized by intestinal bacteria. The body is also able to convert synthetically prepared menaquinone (Menadione) and a number of water-soluble analogs to the biologically active vitamin K_1 (see Figure 26.5). Dietary requirements are measured in terms of micrograms of vitamin K_1 with the RDA for adults being in the range of 70–140 μg.

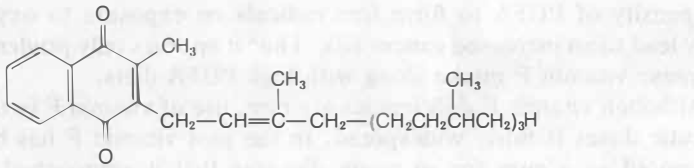
Vitamin K_1 has been shown to be required for the conversion of several clotting factors and prothrombin to the active state. The mechanism of this action has been most clearly delineated for pro-

CLIN. CORR. 26.3 NUTRITIONAL CONSIDERATIONS FOR THE NEWBORN

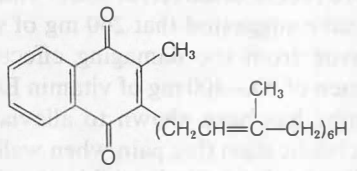
Newborn infants are at special nutritional risk. In the first place, this is a period of very rapid growth, and needs for many nutrients are high. Some micronutrients (such as vitamins E and K) do not cross the placental membrane well and tissue stores are low in the newborn infant. The gastrointestinal tract may not be fully developed, leading to malabsorption problems (particularly with respect to the fat-soluble vitamins). The gastrointestinal tract is also sterile at birth, and the intestinal flora that normally provide significant amounts of certain vitamins take several days to become established. If the infant is born prematurely, the nutritional risk is slightly greater, since the gastrointestinal tract will be less well developed and the tissue stores will be less.

The most serious nutritional complication of newborns appears to be hemorrhagic disease of the newborn. Newborn infants, especially premature infants, have low tissue stores of vitamin K and lack the intestinal flora necessary to synthesize the vitamin. Breast milk is also a relatively poor source of vitamin K. Approximately one out of 400 live births shows some signs of hemorrhagic disease. At birth 1 mg of the vitamin is usually sufficient to prevent hemorrhagic disease.

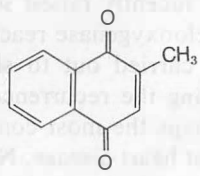
Iron is another potential problem. Most newborn infants are born with sufficient reserves of iron to last 3–4 months (although premature infants are born with smaller reserves). Since iron is present in low amounts in both cow's milk and breast milk, iron supplementation is usually



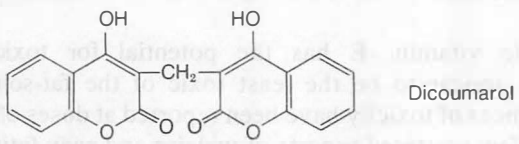
Vitamin K_1 (green leafy vegetables)



Vitamin K_2 (intestinal bacteria)



Vitamin K_3 (menadione)



Dicoumarol

Figure 26.5
Structures of vitamin K and related compounds.

thrombin. Prothrombin is apparently synthesized in an inactive precursor form. Conversion to the active form requires a vitamin K-dependent carboxylation of certain glutamic acid residues to γ-carboxyglutamic acid (Figure 26.6). The γ-carboxyglutamic acid residues are good chelators and allow prothrombin to bind calcium. The prothrombin–Ca^{2+} complex in turn binds to the phospholipid membrane, where proteolytic conversion to thrombin can occur in vivo. The mechanism of the carboxylation reaction has not been fully clarified, but appears to involve the intermediate formation of a

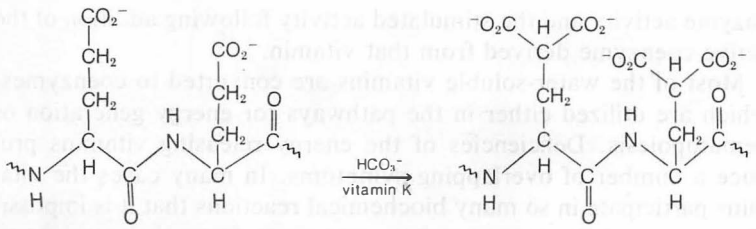


Figure 26.6
Function of vitamin K.

2,3-epoxide derivative of vitamin K. Dicumarol, a naturally occurring anticoagulant, may inhibit the reductase which converts the epoxide back to the active vitamin.

The only known symptom of vitamin K deficiency in man is increased coagulation time. Since vitamin K is relatively abundant in the diet and synthesized in the intestine, deficiencies are very rare. The most common deficiency is seen in newborn infants (Clin. Corr. 26.3). Vitamin K deficiency is also seen in patients with obstructive jaundice and other diseases leading to severe fat malabsorption and patients on long-term antibiotic therapy (which may destroy vitamin K-synthesizing organisms in the intestine). Finally, vitamin K deficiency is occasionally seen in the elderly, who are prone to poor liver function (reducing prothrombin synthesis) and fat malabsorption. Certainly vitamin K deficiency should be suspected in any patient demonstrating easy bruising and prolonged clotting time.

26.4 WATER-SOLUBLE VITAMINS

The water-soluble vitamins differ from the fat-soluble vitamins in several important aspects. In the first place, most of these compounds are readily excreted once their concentration surpasses the renal threshold. Thus toxicities are very rare. It is popular to speak of these vitamins as "not being stored by the body." While that is not quite accurate, the metabolic stores are quite labile and depletion can often occur in a matter of weeks or months. Since the water-soluble vitamins are coenzymes for many common biochemical reactions, it is often possible to assay vitamin status by measuring one or more enzyme activities in isolated red blood cells. These assays are especially useful if one measures both the endogenous

begun at a relatively early age by the introduction of iron-fortified cereal. Vitamin D levels are also considered to be somewhat low in breast milk and supplementation with vitamin D is usually recommended. However, some recent studies have suggested that the iron in breast milk is present in a form that is particularly well utilized by the infant and that earlier studies probably underestimated the amount of vitamin D available in breast milk. Other vitamins and minerals appear to be present in adequate amounts in breast milk as long as the mother is getting a good diet.

Artificial formulas present some special nutritional concerns. The high temperatures necessary for pasteurization and sterilization can lead to loss of C and B_6. There have been instances in which B_6-deficient formulas have reached the marketplace, resulting in a severe enough deficiency to cause convulsions. Vitamin E levels have occasionally caused problems in artificial formulas. While breast milk contains sufficient quantities of vitamin E, cow's milk does not. This can occasionally lead to a hemolytic anemia, especially if the formulas are also high in iron and polyunsaturated fatty acids. However, manufacturers appear to have learned from past mistakes, and most artificial formulas do contain enough of these three nutrients now.

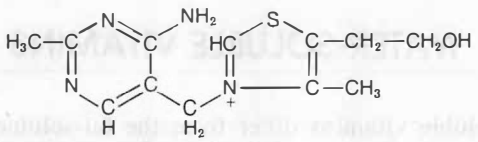
In summary, most infants are provided with supplemental vitamin K at birth to prevent hemorrhagic disease. Breast-fed infants are usually provided with supplemental vitamin D, with iron being introduced along with solid foods. Bottle-fed infants are provided with supplemental iron. The formulas themselves are usually fortified with vitamins D, E, C, and B_6.

enzyme activity and the stimulated activity following addition of the active coenzyme derived from that vitamin.

Most of the water-soluble vitamins are converted to coenzymes, which are utilized either in the pathways for energy generation or hematopoiesis. Deficiencies of the energy releasing vitamins produce a number of overlapping symptoms. In many cases the vitamins participate in so many biochemical reactions that it is impossible to pinpoint the exact biochemical cause of any given symptom. However, it is possible to generalize that because of the central role these vitamins play in energy metabolism, deficiencies show up first in rapidly growing tissues. Typical symptoms include dermatitis, glossitis (swelling and reddening of the tongue), cheilitis at the corners of the lips, and diarrhea. In many cases nervous tissue is also involved due to its high energy demand or specific effects of the vitamin. Some of the common neurological symptoms include peripheral neuropathy (tingling of nerves at the extremities), depression, mental confusion, lack of motor coordination, and malaise. In some cases demyelination and degeneration of nervous tissue also take place. These deficiency symptoms are so common and overlapping that they can be considered as properties of the energy-releasing vitamins as a class, rather than being specific for any one.

26.5 ENERGY-RELEASING WATER-SOLUBLE VITAMINS

Thiamine (Vitamin B_1)



Thiamine

Thiamine is rapidly converted to the coenzyme thiamine pyrophosphate (TPP), which is required for the key reactions catalyzed by pyruvate and α-ketoglutarate dehydrogenases (see Figure 26.7). Thus, the cellular capacity for energy generation is severely compromised in thiamine deficiency. TPP is also required for the transketolase of the pentose phosphate pathway. While the pentose phosphate pathway is not quantitatively important in terms of energy generation, it is the sole source of ribose for the synthesis of

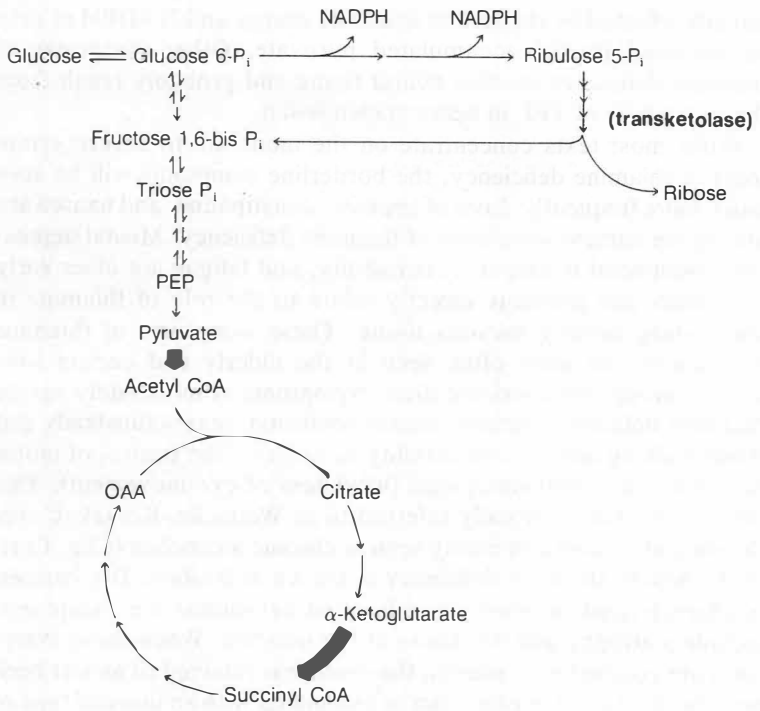


Figure 26.7
Summary of important reactions involving thiamine pyrophosphate. *The reactions involving thiamine pyrophosphate are indicated in boldface type or heavy arrows.*

nucleic acid precursors and the major source of NADPH for fatty acid biosynthesis and other biosynthetic pathways. The red blood cell transketolase is also the enzyme most commonly used for measuring thiamine status in the body. Finally, TPP appears to play an important role in the transmission of nerve impulses. TPP (or a related metabolite, thiamine triphosphate) is localized in peripheral nerve membranes. It appears to be required for acetylcholine synthesis and may also be required for ion translocation reactions in stimulated neural tissue.

Although the biochemical reactions involving TPP are fairly well characterized, it is not clear how these biochemical lesions result in the symptoms of thiamine deficiency. The pyruvate dehydrogenase and transketolase reactions are the most sensitive to thiamine levels. Thus thiamine deficiency appears to selectively inhibit carbohydrate metabolism, causing a buildup of pyruvate. The cells may be

CLIN. CORR. 26.4 NUTRITIONAL CONSIDERATIONS IN THE ALCOHOLIC

Chronic alcoholics run considerable risk of nutritional deficiencies. The most common problems are neurologic symptoms associated with thiamine or pyridoxine deficiencies and hematological problems associated with folate or pyridoxine deficiencies. The deficiencies seen with alcoholics are not necessarily due to poor diet alone, although it is often a strong contributing factor. Alcohol causes pathological alterations of the gastrointestinal tract, which often directly interfere with absorption of certain nutrients. Also the liver is often the most important site of activation and storage of many vitamins. The severe liver damage associated with chronic alcoholism often leads to impaired utilization and storage of certain nutrients.

Up to 40% of hospitalized alcoholics are estimated to have megaloblastic erythropoiesis due to folate deficiency. Alcohol appears to directly interfere with folate absorption and alcoholic cirrhosis impairs both activation and storage of this nutrient. Another 30% of hospitalized alcoholics have sideroblastic anemia or identifiable sideroblasts in erythroid marrow cells characteristic of pyridoxine deficiency. The problem here appears to be an impaired utilization of pyridoxine with much of the pyridoxine being degraded rather than being activated to pyridoxal phosphate. Some alcoholics also develop a peripheral neuropathy which responds to pyridoxine supplementation.

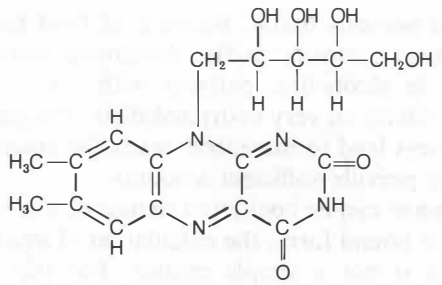
The most common nutritionally related neurological disorder is the Wernicke–Korsakoff Syndrome. The symptoms indirectly affected by the lack of available energy and NADPH or may be poisoned by the accumulated pyruvate. Other symptoms of thiamine deficiency involve neural tissue and probably result from the direct role of TPP in nerve transmission.

While most texts concentrate on the moderate to severe symptoms of thiamine deficiency, the borderline symptoms will be seen much more frequently. Loss of appetite, constipation, and nausea are among the earliest symptoms of thiamine deficiency. Mental depression, peripheral neuropathy, irritability, and fatigue are other early symptoms and probably directly relate to the role of thiamine in maintaining healthy nervous tissue. These symptoms of thiamine deficiencies are most often seen in the elderly and certain low-income groups on restricted diets. Symptoms of moderately severe thiamine deficiency include mental confusion, ataxia (unsteady gait while walking and general inability to achieve fine control of motor functions), and opthalmoplegia (weakness of eye movement). This set of symptoms is usually referred to as Wernicke–Korsakoff syndrome and is most commonly seen in chronic alcoholics (Clin. Corr. 26.4). Severe thiamine deficiency is known as *beriberi*. Dry beriberi is characterized primarily by advanced neuromuscular symptoms, including atrophy and weakness of the muscles. When these symptoms are coupled with edema, the disease is referred to as wet beriberi. Both forms of beriberi can be associated with an unusual type of heart failure characterized by high cardiac output. Beriberi is found primarily in populations relying exclusively on polished rice for food, although cardiac failure is sometimes seen in alcoholics as well.

The thiamine requirement is proportional to the caloric content of the diet and will be in the range of 1.4–1.5 mg for the normal adult. This requirement should be raised somewhat if carbohydrate intake is excessive or if the metabolic rate is elevated (due to fever, trauma, pregnancy, or lactation). Coffee and tea both contain substances that destroy thiamine, but this is not a problem for individuals consuming normal amounts of these beverages. The routine enrichment of cereals has assured that most Americans have an adequate intake of thiamine on a normal mixed diet.

Riboflavin

Riboflavin is converted to the coenzymes flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), both of which are involved in a wide variety of redox reactions. The flavin coenzymes are essential for energy production and cellular respiration. The most characteristic symptoms of riboflavin deficiency are angular cheilitis, glossitis, and scaly dermatitis (especially around the naso-

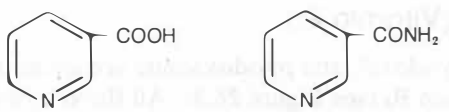


Riboflavin [7,8-dimethyl-10-(1′-D-ribityl)isoalloxazine]

labial folds and scrotal areas). The best flavin-requiring enzyme for assaying riboflavin status appears to be erythrocyte glutathione reductase. The recommended riboflavin intake is 1.2–1.7 mg/day for the normal adult. Foods rich in riboflavin include milk, meat, eggs, and cereal products. Riboflavin deficiencies are quite rare in this country. When riboflavin deficiency does occur, it is usually seen in chronic alcoholics.

Niacin

Niacin is not a vitamin in the strictest sense of the word, since the body is capable of making niacin from tryptophan. However, the conversion of tryptophan to niacin is relatively inefficient (60 mg of tryptophan are required for the production of 1 mg of niacin) and occurs only after all of the other body requirements for tryptophan (protein synthesis and energy production) have been met. Since the synthesis of niacin requires thiamine, pyridoxine, and riboflavin, it is very inefficient on a marginal diet. Thus, in practical terms, most individuals require dietary sources of both tryptophan and niacin. Niacin (nicotinic acid) and niacinamide (nicotinamide) are both converted to the ubiquitous oxidation–reduction coenzymes NAD and NADP in the body.



Niacin Nicotinamide

Borderline deficiencies of niacin are first seen as a glossitis of the tongue, somewhat similar to riboflavin deficiency. Pronounced deficiencies lead to pellagra which is characterized by the three D's: dermatitis, diarrhea, and dementia. The dermatitis is characteristic in that it is usually seen only in skin areas exposed to sunlight and is symmetric. The neurologic symptoms are associated with actual

clude mental disturbances, ataxia (unsteady gait and lack of fine motor coordination), and weakness of eye movements. Congestive heart failure similar to that seen with beriberi is also seen in a small number of these patients. While this syndrome may only account for 1–3% of alcohol-related neurologic disorders, the response to supplemental thiamine is so dramatic that it is usually worth consideration. The thiamine deficiency appears to arise from impaired absorption and utilization as well as from an increased need of the vitamin for acetaldehyde metabolism.

While those are the most common nutritional deficiencies associated with alcoholism, deficiencies of almost any of the water-soluble vitamins can occur and cases of alcoholic scurvy and pellagra are occasionally reported. Alcoholic patients do have decreased bone density and an increased incidence of osteoporosis. This probably relates to the lack of the 25-hydroxylation step in the liver as well as an increased rate of metabolism of vitamin D to inactive products by an activated cytochrome P_{450} system. Dietary calcium intake is also often poor. In fact, alcoholics generally have decreased serum levels of zinc, calcium, and magnesium due to poor dietary intake and increased urinary losses. Iron-deficiency anemia is very rare unless there is gastrointestinal bleeding or chronic infection. In fact, excess iron is a more common problem with alcoholics. Many alcoholic beverages contain relatively high iron levels, and alcohol appears to enhance iron absorption.

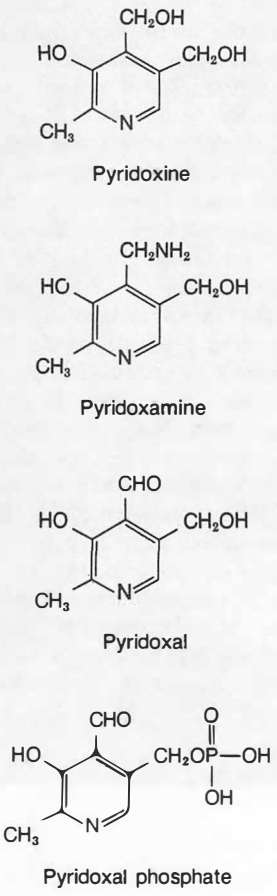


Figure 26.8
Structures of vitamin B_6.

degeneration of nervous tissue. Because of food fortification, pellagra is a medical curiosity in the developed world. Today it is primarily seen in alcoholics, patients with severe malabsorption problems, and elderly on very restricted diets. Pregnancy, lactation, and chronic illness lead to increased needs for niacin, but a varied diet will usually provide sufficient amounts.

Since tryptophan can be converted to niacin, and niacin itself can exist in a free or bound form, the calculation of available niacin for any given food is not a simple matter. For this reason, niacin requirements are expressed in terms of niacin equivalents (1 niacin equiv = 1 mg free niacin). The current recommendation of the Food and Nutrition Board for a normal adult is 13–19 niacin equivalents (N.E.) per day. The richest food sources of niacin are meats, peanuts and other legumes, and enriched cereals.

When nicotinic acid (but not nicotinamide) is used in pharmacologic doses (2–4 g/day) it appears to cause a number of metabolic effects in the body not related to its normal function as a vitamin. For example, vasodilation (flushing) is a very immediate reaction. Over the longer term there is a decreased mobilization of fatty acids from adipose tissue, a marked decrease in circulating cholesterol and lipoproteins (especially LDL), and an elevation of serum glucose and uric acid. These effects can be explained in part by an effect of nicotinic acid on cAMP levels. While the cholesterol lowering effects of nicotinic acid may be desirable in certain controlled clinical situations, there are potential side effects of pharmacologic doses of this vitamin. The reduced mobilization of fatty acids from adipose tissue causes depletion of the glycogen and fat reserves in skeletal and cardiac muscle. The tendency toward elevated glucose and uric acid could cause problems if someone is borderline for diabetes or gout. Finally, continued use of nicotinic acid in those doses is sometimes associated with elevated serum enzymes suggestive of liver damage.

Pyridoxine (Vitamin B_6)

Pyridoxine, pyridoxal, and pyridoxamine are all naturally occurring forms of vitamin B_6 (see Figure 26.8). All three forms are efficiently converted by the body to pyridoxal phosphate which is required for the synthesis, catabolism, and interconversion of amino acids. The role of pyridoxal phosphate in amino acid metabolism has been discussed previously (Chapter 11) and will not be considered here. While pyridoxal phosphate dependent reactions are legion, there are a few instances in which the biochemical lesion seems to be directly associated with the symptoms of B_6 deficiency. Some of these more important reactions are summarized in Figure 26.9. Obviously,

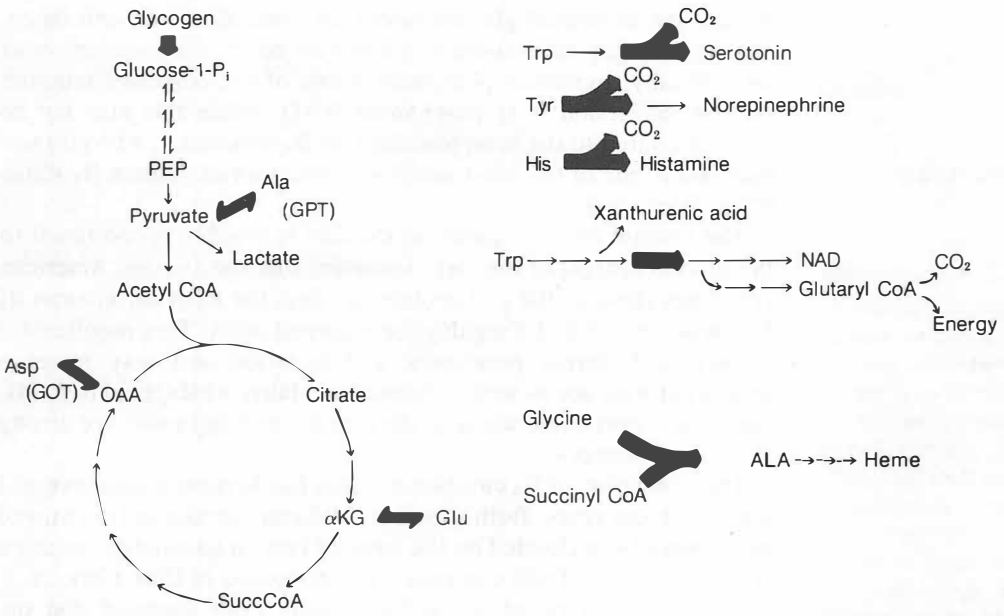


Figure 26.9
Some important metabolic roles of pyridoxal phosphate.
The reactions requiring pyridoxal phosphate are indicated with heavy arrows. ALA = δ-aminolevulinic acid; αKG = α-ketoglutarate; GPT = glutamate pyruvate transaminase; and GOT = glutamate oxalacetate transaminase.

pyridoxal phosphate is essential for energy production from amino acids and can be considered an energy releasing vitamin. Thus some of the symptoms of severe B_6 deficiency are similar to those of the other energy-releasing vitamins. Pyridoxal phosphate is also required for the synthesis of the neurotransmitters serotonin and norepinephrine and appears to be required for the synthesis of the sphingolipids necessary for myelin formation. These effects are thought to explain the irritability, nervousness, and depression seen with mild deficiencies and the peripheral neuropathy and convulsions observed with severe deficiencies. Pyridoxal phosphate is required for the synthesis of δ-aminolevulinic acid, a precursor of heme. B_6 deficiencies occasionally cause sideroblastic anemia, which is characteristically a microcytic anemia seen in the presence of high serum iron. Pyridoxal phosphate is also an essential component of the enzyme glycogen phosphorylase. It is covalently linked to a lysine residue and stabilizes the enzyme. This role of B_6 appears to

CLIN. CORR. 26.5 VITAMIN B_6 REQUIREMENTS FOR USERS OF ORAL CONTRACEPTIVES

The controversy over vitamin B_6 requirements for users of oral contraceptives best illustrates the potential problems associated with biochemical assays. For years, one of the most common assays for vitamin B_6 status has been the tryptophan load assay. This assay is based on the observation that when tissue pyridoxal phosphate levels are low, the normal catabolism of tryptophan is impaired and most of the tryptophan is catabolized by a minor pathway leading to synthesis of xanthurenic acid (Figure 26.9). Under many conditions, the amount of xanthurenic acid recovered in a 24-h urine sample following ingestion of a fixed amount of tryptophan is a valid indicator of vitamin B_6 status. When the tryptophan load test was used to assess the vitamin B_6 status of oral contraceptive users, however, alarming reports started appearing in the literature. Not only did oral contraceptive use increase the excretion of xanthurenic acid considerably, but the amount of pyridoxine hydrochloride needed to return xanthurenic acid excretion to normal was 20 mg/day. This amounts to 10 times the RDA and almost 20 times the level required to maintain normal B_6 status in control groups. As might be expected, this observation received much popular attention in spite of the fact that most classical symptoms of vitamin B_6 deficiency were not observed in oral contraceptive users.

More recent studies using other measures of vitamin B_6 have painted a slightly different picture. For example, erythrocyte glutamate pyruvate transaminase (EGPT) and erythrocyte glutamate oxal-

explain the decreased glucose tolerance associated with deficiency, although B_6 may have some direct effects on insulin metabolism as well. Finally, pyridoxal phosphate is one of the cofactors required for the conversion of tryptophan to NAD. While this may not be directly related to the symptomology of B_6 deficiency, a tryptophan load test is one of the most sensitive indicators of vitamin B_6 status (Clin. Corr. 26.5).

The amount of B_6 required in the diet is roughly proportional to the protein content of the diet. Assuming that the average American consumes close to 100 g of protein per day, the RDA for vitamin B_6 has been set at 1.8–2.2 mg/day for a normal adult. This requirement is increased during pregnancy and lactation and may increase somewhat with age as well. Vitamin B_6 is fairly widespread in foods, but meat, vegetables, whole-grain cereals, and egg yolks are among the richest sources.

The evaluation of B_6 nutritional status has become a controversial topic in recent years. Both questions of dietary intake and nutritional needs have been clouded by the issue of how to adequately measure B_6 status. Some of this controversy is discussed in Clin. Corr. 26.5. In terms of dietary intake, it has usually been assumed that the average American diet is adequate in B_6 and it is not routinely added to flour and other fortified foods. However, some recent nutritional surveys have cast doubt on that assumption. In several instances, a significant fraction of the survey population was found to consume less than two-thirds of the RDA for B_6. In none of these surveys were clinical symptoms of B_6 deficiency observed. While dietary intake of B_6 may be marginal for many individuals, this does not appear to cause adverse affects unless coupled with increased demand. Pregnancy, for example, is usually considered to increase the needs for B_6. The usual recommendation is for an additional intake of 0.5 mg/day.

There are, however, a few instances where B_6 deficiencies are clear-cut and noncontroversial. For example, newborn infants rapidly develop symptoms of hyperirritability and convulsive seizures when fed milk or formulas containing less than 50 μg of vitamin B_6/liter. Also the drug isoniazid (isonicotinic acid hydrazide), which is commonly used in the treatment of tuberculosis reacts with pyridoxal or pyridoxal phosphate to form a hydrazone derivative, which inhibits pyridoxal phosphate-containing enzymes. Patients on long-term isoniazid treatment develop a peripheral neuropathy, which responds well to B_6 therapy. Finally, penicillamine (β-dimethylcysteine), which is used in the treatment of patients with Wilson's disease, cystinuria, and rheumatoid arthritis, reacts with pyridoxal phosphate to form an inactive thiazolidine derivative. Patients treated with penicillamine occasionally develop convulsions, which can be prevented by B_6 supplementation.

Other Energy-Releasing Vitamins

Pantothenic acid is an essential component of coenzyme A (CoA) and acyl carrier protein (ACP) and thus is required for the metabolism of all fat, protein, and carbohydrate via the citric acid cycle. In short, more than 70 enzymes have been described to date which utilize CoA or ACP derivatives. In view of the importance of these reactions, one would expect pantothenic acid deficiencies to be a serious concern in man. However, this does not appear to be the case and the reasons are essentially twofold: (1) pantothenic acid is very widespread in natural foods–probably reflecting its widespread metabolic role, and (2) most symptoms of pantothenic acid deficiency are vague and mimic those of other B vitamin deficiencies.

Biotin is the prosthetic group for a number of carboxylation reactions, the most notable being pyruvate carboxylase (needed for synthesis of oxalacetate for gluconeogenesis and replenishment of the citric acid cycle) and acetyl CoA carboxylase (fatty acid biosynthesis). Biotin is found in peanuts, chocolate, and eggs and is usually synthesized in more than adequate amounts by intestinal bacteria. Biotin deficiency is generally seen only following long-term antibiotic therapy or excessive consumption of raw egg white. The raw egg white contains a protein, avidin, which binds biotin in a nondigestible form. However, in humans raw egg white must comprise 30% of the caloric intake (approximately 20 egg whites/day) to precipitate a biotin deficiency.

26.6 HEMATOPOIETIC WATER-SOLUBLE VITAMINS

Folic Acid (Folacin)

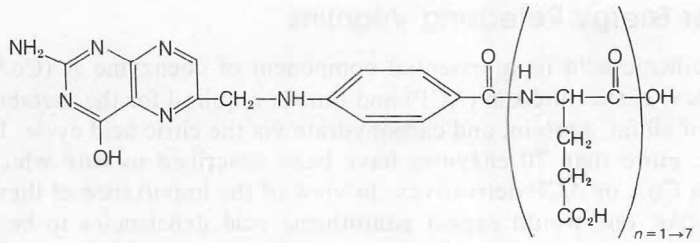
The simplest form of folic acid is monopteroylglutamic acid. However, folic acid usually occurs as polyglutamate derivatives with from 2 to 7 glutamic acid residues (Figure 26.10). These compounds are taken up by intestinal mucosal cells and the extra glutamate residues removed by deconjugase, a lysosomal enzyme. The free folic acid is then reduced to tetrahydrofolate by the enzyme dihydrofolate reductase and circulated in the plasma primarily as the N^5-methyl derivative of free tetrahydrofolate (Figure 26.10). However, inside the cells, tetrahydrofolates are found primarily as polyglutamate derivatives, and these appear to be the biologically most potent forms. Folic acid is also stored as a polyglutamate derivative of tetrahydrofolate in the liver.

acetate transaminase (EGOT) are both pyridoxal phosphate-containing enzymes. One can also assess vitamin B_6 status by measuring the endogenous activity of these enzymes and the degree of stimulation by added pyridoxal phosphate. These types of assays show a much smaller difference between nonusers and users of oral contraceptives. The minimum level of pyridoxine hydrochloride needed to maintain normal vitamin B_6 status as measured by these assays was only 2.0 mg/day, which is equal to the RDA and only 1.5 to 2.0 times greater than that needed by nonusers.

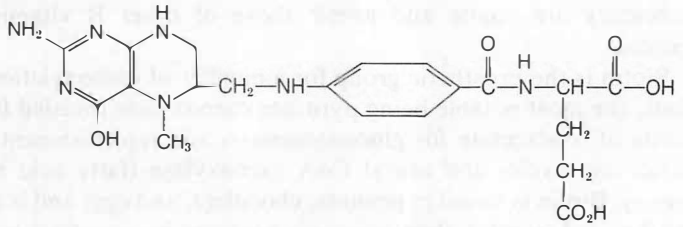
Why the large discrepancy? For one thing, it must be kept in mind that for any sort of assay such as the tryptophan load test, what you are actually measuring is the activity of a pyridoxal phosphate containing enzyme *relative* to the overall need for that enzyme. Any unrelated event which increased the overall rate of tryptophan catabolism via the pathway in Figure 26.9, would increase the need for the pyridoxal phosphate-dependent enzyme(s) in that pathway. For example, it is known that the levels of tryptophan dioxygenase (the first enzyme of the pathway) can be increased by hormonal stimulus (for example by hydrocortisone). It seems probable then, that the estrogens in oral contraceptives may also increase the levels of tryptophan dioxygenase, thus creating an increased pyridoxine requirement for tryptophan metabolism without necessarily affecting pyridoxine requirements for other metabolic processes in the body.

Does this mean that vitamin B_6 status is of no concern to users of oral contraceptives? Oral contraceptives do appear to increase vitamin B_6 requirements slightly. Several dietary surveys have shown that a significant percentage of women in the 18- to 24-year age group consume diets containing less than 1.3 mg pyridoxine/day. If these women are also using oral contraceptives, they are at some increased

risk for developing a borderline deficiency. Furthermore, there are documented cases of depression and decreased glucose tolerance in oral contraceptive users, which responded to pyridoxine supplementation. While these cases are the exception rather than the rule, they do demonstrate that poor diet plus the slightly increased demands for B_6 due to oral contraceptive use can be expected to lead to minor deficiencies in some individuals. While the tryptophan load test was clearly misleading in a quantitative sense, it did alert the medical community to a previously unsuspected nutritional risk.



Folic acid



N^5-Methyltetrahydrofolate

Figure 26.10
Structure of folic acid and N^5-methyltetrahydrofolate.

Various 1-carbon tetrahydrofolate derivatives are used in biosynthetic reactions (see Figure 26.11). They are required, for example, in the synthesis of choline, serine, glycine, methionine, purines, and dTMP. Since adequate amounts of choline and the amino acids can usually be obtained from the diet, the participation of folates in purine and dTMP synthesis appears to be metabolically most significant. However, under some conditions the folate-dependent conversion of homocysteine to methionine can make a significant contribution to the available methionine pool. Methionine, of course, is converted to *S*-adenosylmethionine, which is also used in a number of biologically important methylation reactions.

The most pronounced effect of folate deficiency is inhibition of DNA synthesis due to decreased availability of purines and dTMP. This leads to an arrest of cells in S phase and a characteristic "megaloblastic" change in the size and shape of the nuclei of rapidly dividing cells. The block in DNA synthesis also slows down the maturation of red blood cells, causing production of abnormally large "macrocytic" red blood cells with fragile membranes. The

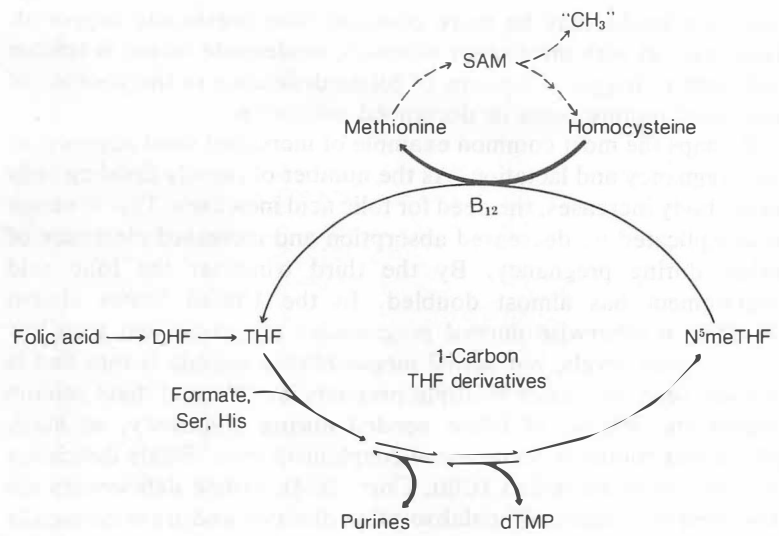


Figure 26.11
Metabolism of folic acid.
The metabolic interconversions of folic acid and its derivatives are indicated with light arrows. Reactions or pathways requiring folate derivatives are indicated with heavy arrows. Other related pathways are indicated with dotted arrows. DHF = dihydrofolate, THF = tetrahydrofolate, N^5meTHF = N^5-methyltetrahydrofolate, dTMP = deoxythymidylic acid, SAM = S-adenosylmethionine.

rapid hemolysis of these macrocytes leads to a hemolytic anemia. Thus a macrocytic anemia associated with megaloblastic changes in the bone marrow is fairly characteristic of folate deficiency.

There is some uncertainty over the incidence of folate deficiencies. Folates occur very widely in foods, especially meats and a variety of fresh fruits and vegetables. However, these folates occur in a variety of different forms, which are utilized with varying efficiencies. Assessment of nutritional status is further complicated by the use of different biochemical methods to measure folate levels. For example, the serum folate level decreases rapidly on restricted folate intake and is a very sensitive indicator of folate status. However, levels of polyglutamate folate derivatives in red blood cells decrease much more slowly and are more indicative of tissue levels. Symptoms of folate deficiency appear only as tissue folates are depleted. There can be many causes of folate deficiency, including inadequate intake, impaired absorption, increased demands, and impaired metabolism. Some dietary surveys have suggested that in-

CLIN. CORR. 26.6 ANTICONVULSANT DRUGS AND VITAMIN REQUIREMENTS

Anticonvulsant drugs such as phenobarbital or diphenylhydantoin (DPH) present an excellent example of the type of drug–nutrient interactions which are of concern to the physician. Anticonvulsants inhibit the conversion of vitamin D to the 25-hydroxy derivative in the liver and enhance its conversion to inactive metabolites. Whereas children and adults on these drugs seldom develop rickets or osteomalacia, as many as 65% of those on long-term therapy will have abnormally low serum calcium and phosphorus and abnormally high serum alkaline phosphatase. Some bone loss is usually observed in these cases. The amount of supplemental vitamin D required to correct this problem is one and a half to two and a half times the RDA. Anticonvulsants also tend to increase needs for vitamin K, leading to an increased incidence of hemorrhagic disease in infants born to mothers on anticonvulsants.

Anticonvulsants also seem to increase the need for folic acid and B_6. Low serum folate levels are seen in 75% of patients on anticonvulsants and megaloblastic anemia may occur in as many as 50% without supplementation. By biochemical parameters, 30–60% of the children on anticonvulsants exhibit some form of B_6 deficiency. However, clinical symptoms of B_6 deficiency are rarely seen; 1–5 mg of folic acid and 10 mg of vitamin B_6 appear to be sufficient for most patients on anticonvulsants. Since folates may speed up the metabolism of some anticonvulsants, it is important that excess folic acid not be given.

adequate intake may be more common than previously supposed. However, as with most other vitamins, inadequate intake is seldom sufficient to trigger symptoms of folate deficiency in the absence of increased requirements or decreased utilization.

Perhaps the most common example of increased need occurs during pregnancy and lactation. As the number of rapidly dividing cells in the body increases, the need for folic acid increases. This situation is complicated by decreased absorption and increased clearance of folate during pregnancy. By the third trimester the folic acid requirement has almost doubled. In the United States almost 20–25% of otherwise normal pregnancies are associated with low serum folate levels, but actual megaloblastic anemia is rare and is usually seen only after multiple pregnancies. Normal diets seldom supply the 800 μg of folate needed during pregnancy, so many physicians routinely recommend supplementation. Folate deficiency is common in alcoholics (Clin. Corr. 26.4). Folate deficiencies are also seen in a number of malabsorption diseases and are occasionally seen in the elderly, due to a combination of poor dietary habits and poor absorption.

There are a number of drugs that also directly interfere with folate metabolism. Anticonvulsants and oral contraceptives interfere with folate absorption (Clin. Corr. 26.6). Oral contraceptives and estrogens also appear to interfere with folate metabolism in their target tissue. Long-term use of any of these drugs can lead to folate deficiencies unless adequate supplementation is provided. For example, 20% of patients using oral contraceptives develop megaloblastic changes in the cervicovaginal epithelium, and 20–30% show low serum folate levels.

Vitamin B_{12} (Cobalamine)

Pernicious anemia, a megaloblastic anemia associated with neurological deterioration, was invariably fatal until 1926 when liver extracts were shown to be curative. Subsequent work showed the need for both an extrinsic factor present in liver and an intrinsic factor produced by the body. Vitamin B_{12} was the extrinsic factor. Chemically, vitamin B_{12} consists of cobalt in a coordination state of six, coordinated in 4 positions by a tetrapyrrol (or corrin) ring, in one position by a benzimidazole nitrogen, and in the sixth position by one of several different ligands (Figure 26.12). The crystalline forms of B_{12} used in supplementation are usually hydroxycobalamine or cyanocobalamine. In foods B_{12} usually occurs bound to protein in the methyl or 5′-deoxyadenosyl forms. To be utilized the B_{12} must first be removed from the protein by acid hydrolysis in the stomach or trypsin digestion in the intestine. It then must combine with "in-

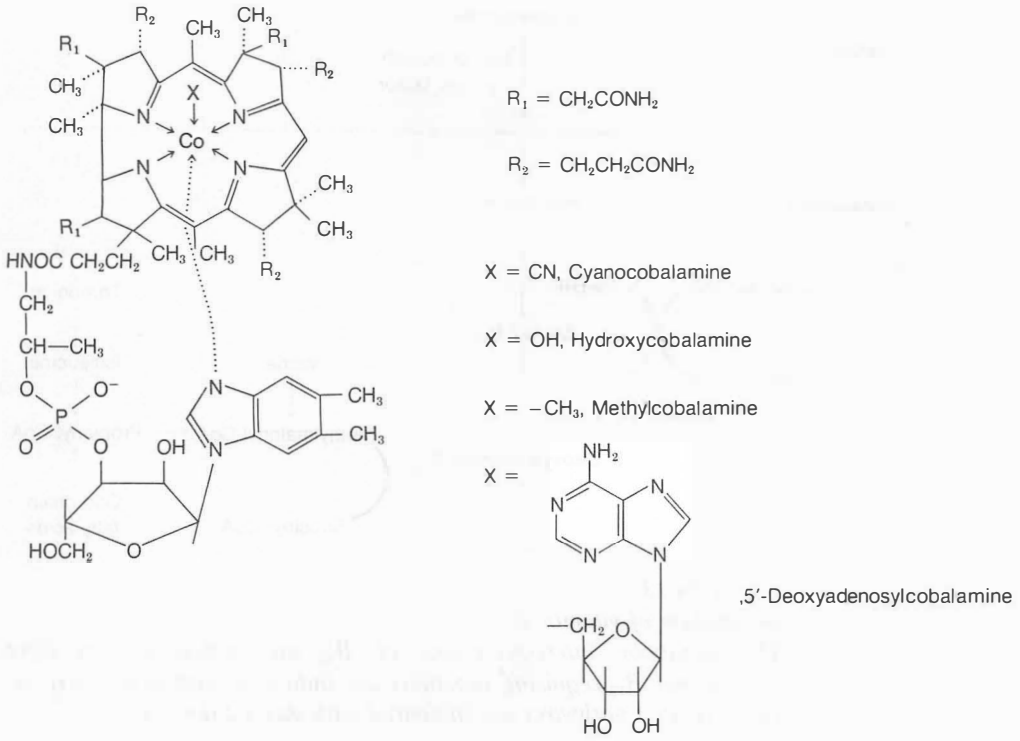


Figure 26.12
Structure of vitamin B_{12} (cobalamine).

trinsic factor," a protein secreted by the stomach, which carries it to the ileum for absorption.

In man there are two major symptoms of B_{12} deficiency (hematopoietic and neurological), and only two biochemical reactions in which B_{12} is known to participate. Thus it is very tempting to speculate on exact cause and effect mechanisms. The methyl derivative of B_{12} is required for the conversion of homocysteine to methionine and the 5-deoxyadenosyl derivative is required for the methylmalonyl CoA mutase reaction (methylmalonyl CoA ⟶ succinyl CoA), which is a key step in the catabolism of some branched-chain amino acids. The neurologic disorders seen in B_{12} deficiency are due to progressive demyelination of nervous tissue. It has been proposed that the methylmalonyl CoA which accumulates interferes with myelin sheath formation in two ways. (1) Methylmalonyl CoA is a competitive inhibitor of malonyl CoA in fatty acid biosynthesis. Since the myelin sheath is continually turning over, any se-

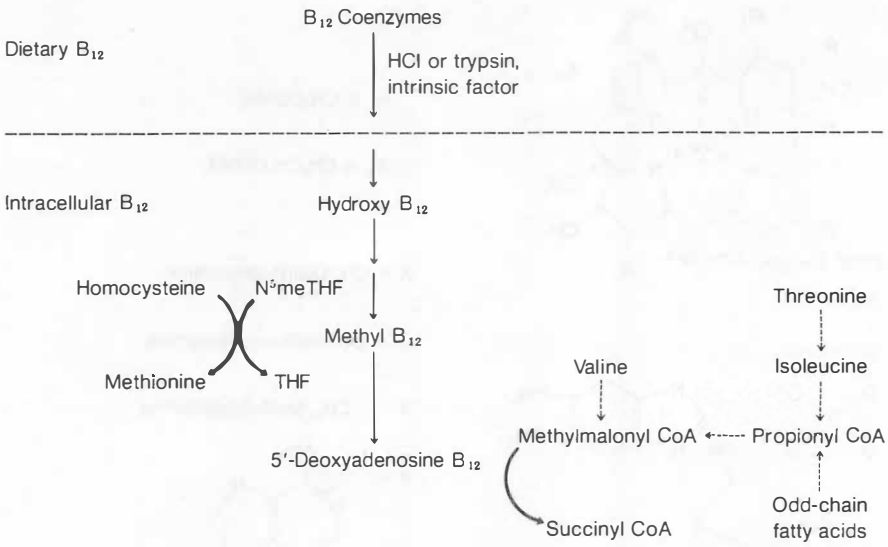


Figure 26.13
Metabolism of vitamin B_{12}.
The metabolic interconversions of B_{12} are indicated with light arrows and B_{12} requiring reactions are indicated with heavy arrows. Other related pathways are indicated with dashed arrows.

vere inhibition of fatty acid biosynthesis will lead to its eventual degeneration. (2) In the residual fatty acid synthesis that does occur, methylmalonyl CoA can substitute for malonyl CoA in the reaction sequence, leading to branched-chain fatty acids, which might disrupt normal membrane structure. There is some evidence supporting both mechanisms.

The megaloblastic anemia associated with B_{12} deficiency is thought to be due to the effect of B_{12} on folate metabolism. The B_{12}-dependent homocysteine to methionine conversion (homocysteine + N^5-methyl THF $\longrightarrow$ methionine + THF) appears to be the only major pathway by which N^5-methyltetrahydrofolate can return to the tetrahydrofolate pool. Thus in B_{12} deficiency there is a buildup of N^5-methyltetrahydrofolate and a deficiency of the tetrahydrofolate derivatives needed for purine and thymine biosynthesis. Essentially all of the folate becomes "trapped" as the N^5-methyl derivative. B_{12} also may be required for uptake of folate by cells and for its conversion to the biologically more active polyglutamate forms. High levels of supplemental folate can overcome the megaloblastic anemia associated with B_{12} deficiencies but not the

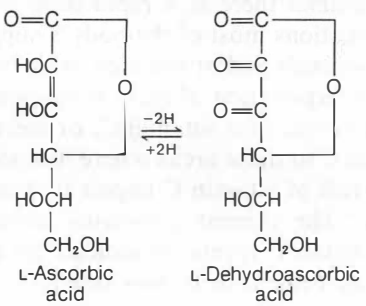
neurological problems. Hence caution must be utilized in using folate to treat megaloblastic anemia.

Vitamin B_{12} is widespread in foods of animal origin, especially meats. Furthermore, the liver stores up to a 6-year supply of vitamin B_{12}. Thus, deficiencies of B_{12} are extremely rare. They are occasionally seen in older people due to insufficient production of intrinsic factor and/or HCl in the stomach. B_{12} deficiency can also be seen in patients with severe malabsorption diseases and in long-term vegetarians.

26.7 OTHER WATER-SOLUBLE VITAMINS

Ascorbic Acid

Vitamin C or ascorbic acid is a 6-carbon compound closely related to glucose. Its main biological role appears to be as a reducing agent in a number of important hydroxylation reactions in the body. For



example, there is clear evidence that ascorbic acid is required for the hyroxylation of lysine and proline in protocollagen. Without the hydroxylation of these amino acids, the protocollagen is unable to properly cross-link into normal collagen fibrils. Thus vitamin C is obviously important for maintenance of normal connective tissue and for wound healing, since the connective tissue is normally laid down first. Vitamin C is also necessary for bone formation, since bone tissue contains an organic matrix containing collagen as well as the inorganic, calcified portion. Finally, collagen appears to be a component of the ground substance surrounding capillary walls.

Since vitamin C is concentrated in the adrenal gland—especially in periods of stress—it has also been postulated to be required for the

hydroxylation reactions involved in the synthesis of some corticosteroids. Ascorbic acid has other important properties as a reducing agent, which appear to be nonenzymatic. For example, it aids in the absorption of iron by reducing it to the ferrous state in the stomach. It spares vitamin A, vitamin E, and some B vitamins by protecting them from oxidation. Finally, it enhances the utilization of folic acid, either by aiding the conversion of folate to tetrahydrofolate or the formation of polyglutamate derivatives of tetrahydrofolate.

Most of the symptoms of vitamin C deficiency can be directly related to its metabolic roles. Symptoms of mild vitamin C deficiency include easy bruising and the formation of petechiae (small, pinpoint hemorrhages in the skin) due to increased capillary fragility. Mild vitamin C deficiencies are also associated with decreased immunocompetence. Scurvy itself is associated with decreased wound healing, osteoporosis, hemorrhaging, and anemia. The osteoporosis results from the inability to maintain the organic matrix of the bone, followed by demineralization. The anemia results from the extensive hemorrhaging coupled with defects in iron absorption and folate activation.

Since vitamin C is readily absorbed, vitamin C deficiencies almost invariably result from poor diet and/or increased need. There is some uncertainty over the need for vitamin C in periods of stress. In severe stress or trauma there is a rapid drop in serum vitamin C levels. In these situations most of the body's supply of vitamin C is mobilized to the adrenals and/or the area of the wound. These facts are clear but the interpretation of them is variable. Does this represent an increased demand for vitamin C, or merely a normal redistribution of vitamin C to those areas where it is needed most? Do the lowered serum levels of vitamin C impair its functions in other tissues in the body? The current consensus appears to be that the lowered serum vitamin C levels do indicate an increased demand, but there is little agreement as to how much.

A similar situation exists with respect to the effect of various drugs on vitamin C status. Smoking has been shown to cause lower serum levels of vitamin C. Aspirin appears to block uptake of vitamin C by platelets. Oral contraceptives and corticosteroids also lower serum levels of vitamin C. While there is no universal agreement as to the seriousness of these effects on vitamin C requirements, the possibility of marginal C deficiencies should be considered with any patient using these drugs over a long period of time, especially if dietary intake is less than optimal.

Of course, the most controversial question surrounding vitamin C is its use in megadoses to prevent and cure the common cold. Ever since this use of vitamin C was first popularized by Linus Pauling in 1970, the issue has generated considerable controversy. However,

some reliable double-blind studies appear to have substantiated the claim in part. The number of colds experienced by vitamin C supplemented groups appeared to be about the same as for control groups, but the severity and duration of the colds were significantly decreased. Thus, while vitamin C does not appear to be useful in preventing the common cold, it does appear to moderate its symptoms. There is no clear indication at present as to how much vitamin C is required to achieve this effect. In the original experiment the control group was eating a balanced diet providing at least 45 mg of ascorbic acid/day while the experimental group was receiving 1–4 g/day. Subsequent experiments have suggested that considerably less vitamin C may achieve the same result. The mechanism by which vitamin C ameliorates the symptoms of the common cold is not known. It has been suggested that vitamin C is required for normal leukocyte function or for synthesis and release of histamine during stress situations. While megadoses of vitamin C are probably no more harmful than the widely used over the counter cold medications, there are some potential side effects of high vitamin C intake which should be considered. For example, oxalate is a major metabolite of ascorbic acid. Thus, high ascorbate intakes could lead to the formation of oxalate kidney stones in predisposed individuals. Pregnant mothers taking megadoses of vitamin C may give birth to infants with abnormally high vitamin C requirements. Earlier suggestions that megadoses of C interfered with B_{12} metabolism have proved to be erroneous.

26.8 MACROMINERALS

Calcium

Calcium is the most abundant mineral in the body. Most of this calcium is in the bone, but the small amount of calcium outside of the bone functions in a number of essential processes. It is required for many enzymes, mediates some hormonal responses, and is essential for blood coagulation. It is also essential for muscle contractability and normal neuromuscular irritability. In fact only a relatively narrow range of serum calcium levels is compatible with life. Since maintenance of constant serum calcium levels is so vital, an elaborate homeostatic control system has evolved. Part of this was discussed earlier in the section on vitamin D metabolism. Low serum calcium stimulates formation of 1,25-dihydroxycholecalciferol, which enhances calcium absorption. If dietary calcium intake is in-

sufficient to maintain serum calcium, 1,25-dihydroxycholecalciferol and parathyroid hormone stimulate bone resorption. Long-term dietary calcium insufficiency, therefore, almost always results in net loss of calcium from the bones.

However, dietary calcium requirements are very difficult to determine due to the existence of other factors that affect availability of calcium. One important factor is vitamin D. Also the phytate (inositol hexaphosphate) found in certain whole-grain foods can reduce calcium uptake. Finally, as discussed in Chapter 25, excess protein in the diet may upset calcium balance by causing more rapid excretion of calcium. Thus calcium balance studies carried out on Peruvian Indians, who have extensive exposure to sunlight and subsist on low protein diets, indicate a need for only 300–400 mg calcium/day. However, calcium balance studies carried out in this country consistently show higher requirements and the RDA has been set at 800 mg/day, with an additional 400-mg allowance for pregnant and lactating women.

The chief symptoms of calcium deficiency are similar to those of vitamin D deficiency, but other symptoms such as muscle cramps are possible with marginal deficiencies. Dietary surveys indicate that a significant portion of certain population groups in this country do not have adequate calcium intake—especially low-income children and adult females. This is of particular interest because these are the population groups with particularly high needs for calcium. For this reason, the U.S. Congress has recently established the WIC Program (Women and Infant Children) to assure adequate protein, calcium, and iron to indigent families with pregnant/lactating mothers or young infants.

Dietary surveys also show that 34–47% of the over-60 population consume less than one-half the RDA for calcium. This is also the age group most at risk of developing osteoporosis, which is characterized by loss of the organic matrix as well as progressive demineralization of the bone. The causes of osteoporosis are multifactorial and largely unknown, but it appears likely that part of the problem has to do with calcium metabolism. The ability to convert vitamin D to the active 1,25-dihydroxy metabolite decreases with age, especially in women past menopause. Recent studies suggest that postmenopausal women may need up to 1,200 mg calcium/day just to maintain calcium balance. There have been suggestions that this elevated calcium requirement, coupled with years of inadequate dietary intake, may be a factor in osteoporosis. There is no good evidence that high dietary calcium alone can prevent osteoporosis, but some studies have suggested that supplemental calcium in the range of 1 g/day may slow the rate of bone loss. This remains a controversial area, with further studies needed.

Other Macrominerals

Phosphorus is a universal constituent of living cells and, for that reason, is almost always present in adequate amounts in the diet. Hypophosphatemia is one symptom of vitamin D deficiency. It can also occasionally be seen following excessive use of antacids containing aluminum hydroxide or calcium carbonate, which form insoluble precipitates with phosphate. Uncontrolled metabolic acidosis can lead to excessive phosphate loss in the urine. The initial symptom of hypophosphatemia is muscle weakness, but eventually a form of rickets can develop.

Magnesium is also ubiquitous in living tissue. It is required for many enzyme activities and for neuromuscular transmission. Diets inadequate in magnesium appear to be rare in this country. However, magnesium deficiency is occasionally observed in conditions of alcoholism, use of certain diuretics, and metabolic acidosis. The main symptoms of magnesium deficiency are weakness, tremors and cardiac arrhythmia. The discovery that heart muscle from patients with myocardial infarctions was low in magnesium lead to the hypothesis that magnesium deficiency might be a predisposing condition for various forms of heart disease. However, there is no evidence that the patients with heart disease had, in fact, consumed diets inadequate in magnesium. The possibility must be considered that a redistribution of tissue magnesium takes place as a result of the heart attack. There is some evidence that supplemental magnesium may help prevent the formation of calcium oxalate stones in the kidney.

26.9 TRACE MINERALS

Iron

Iron metabolism is unique in that it operates largely as a closed system, with iron stores being efficiently reutilized by the body. Not only are iron losses normally minimal (<1 mg/day), but iron absorption is also minimal under the best of conditions. Iron usually occurs in foods in the ferric form bound to protein or organic acids. Before absorption can occur, the iron must be split from these carriers (a process that is facilitated by the acid secretions of the stomach) and reduced to the ferrous form (a process that is enhanced by ascorbic acid). Only 10% of the iron in an average mixed diet is usually absorbed, but the efficiency of absorption can be increased to 30%

CLIN. CORR. 26.7 CALCULATION OF AVAILABLE IRON

The iron RDA for women of childbearing age has been estimated at 18 mg/day, based on a need for 1.5–2.0 mg of absorbed iron and an average absorption of 10%. While this calculation makes meal planning simple enough, it has long been regarded as unsatisfactory. In the first place, the 10% average figure is not very useful, since absorption can vary from 2% to over 30% depending on the food source and need. Also, while it is almost impossible to design a diet containing more than 6 mg iron/1,000 kcal, most American women of childbearing age do not suffer from iron-deficiency anemia. A much more useful method is to calculate the actual amount of dietary iron available for absorption in the diet and compare that to the 1.5–2.0 mg needed. This calculation is based on the following data: (1) 40% of the iron in meat is heme iron and the efficiency of absorption of heme iron is 23%. This does not depend on the presence of other food factors. (2) All other dietary iron is nonheme iron, and absorption is dependent on other food, primarily the presence

Availability	*Consumed in Presence of*	*% Absorption*
Low	(<1 oz of meat or <25 mg C)	3
Medium	(1–3 oz of meat or 25–75 mg C)	5
High	(>3 oz of meat or >75 mg C) or (>1 oz of meat + >25 mg C)	8

by severe iron deficiency. Iron absorption and metabolism have been discussed previously (Chapter 22) and are summarized in Figure 26.14.

Iron, of course, plays a number of important roles in the body. As a component of hemoglobin and myoglobin, it is required for O_2 and CO_2 transport. As a component of cytochromes and nonheme iron proteins, it is required for oxidative phosphorylation. As a compo-

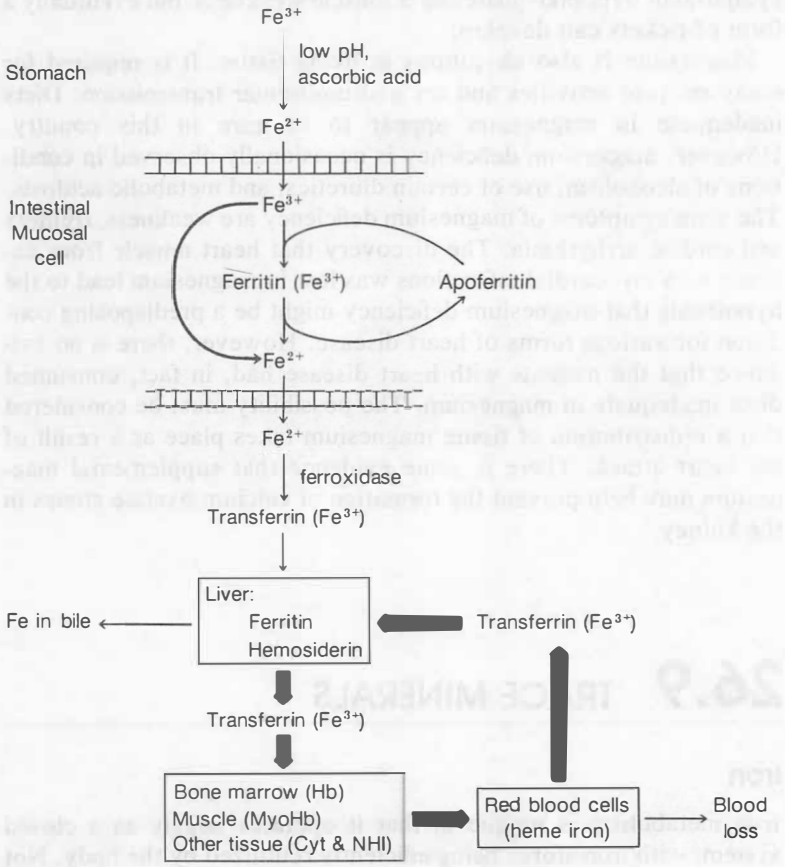


Figure 26.14
Overview of iron metabolism.
This figure reviews some of the features of iron metabolism discussed previously in Chapter 22. The heavy arrows indicate that most of the body's iron is efficiently reutilized by the pathway shown. Hb = hemoglobin, MyoHb = myoglobin, Cyt = cytochromes, and NHI = nonheme iron.

nent of the essential lysosomal enzyme myeloperoxidase, it is required for proper phagocytosis and killing by neutrophils. The best-known symptom of iron deficiency is a microcytic hypochromic anemia. Iron deficiency is also associated with decreased immunocompetence.

Assuming a 10% efficiency of absorption, the Food and Nutrition Board has set a recommended dietary allowance of 10 mg/day for a normal adult male and 18 mg/day for a menstruating female (see Clin. Corr. 26.7). For pregnant and lactating females these allowances are raised to 30–60 mg/day. While 10 mg of iron can easily be obtained from a normal diet, 18 mg is marginal at best and 30–60 mg can almost never be obtained. The best dietary sources of iron are meats, dried legumes, dried fruits, and enriched cereal products.

Iron deficiency anemia has long been considered the most prevalent nutritional disorder in the United States. Young children need enough iron to allow for a continuing increase in blood volume, as do pregnant females. Menstruating females lose iron through blood loss and lactating females through production of lactoferritin. Thus iron deficiency anemia is primarily a problem for these population groups. This is reflected in dietary surveys, which indicate that 95% or more of children and menstruating females are not obtaining adequate iron in their diet. It is also reflected in biochemical measurements of a 10–25% incidence of iron deficiency anemia in this same group. Iron-deficiency anemia is also occasionally a problem with the elderly due to poor dietary intake and increased frequency of achlorhydria.

Because of the widespread nature of iron-deficiency anemia, government programs of nutritional intervention such as the WIC program have emphasized iron rich foods. There has also been discussion of more extensive iron fortification of foods. There is concern among some nutritionists that iron deficiency has been overemphasized (Clin. Corr. 26.8). Since iron excretion is very limited, it is possible to build up toxic levels of iron. The excess iron leads to a condition called hemochromatosis in which iron deposits are found in abnormally high levels in many tissues. This can lead to liver, pancreatic, and cardiac dysfunction as well as pigmentation of the skin. This condition is usually only seen in hemolytic anemias and liver disease, but some concern has been voiced that if iron fortification of foods were to become more widespread, iron overload could become more prevalent.

Iodine

Dietary iodine is very efficiently absorbed and transported to the thyroid gland, where it is stored and used for the synthesis of the

of ascorbic acid and/or some factor(s) present in meat.

Calculations based on this information provide a much more rational estimate of dietary iron. For example, 3 oz of soybeans contain more than twice as much iron as a 3-oz steak. However, at 3% availability, they would provide only 60% as much available iron. By improving availability to 5%, one could obtain an equivalent amount of iron. This adds a new dimension to planning vegetarian diets. Not only should the protein be balanced and include ample quantities of rich iron sources (dried beans and dried fruit), but these meals should be consumed along with fresh fruit and/or fruit juices to provide enough vitamin C for efficient absorption. While 75 mg of vitamin C might be difficult to obtain in a single meal without supplementation, many fresh fruits and fruit juices will supply 50 mg.

These calculations are also useful in planning diets containing iron-fortified food products. While $FeSO_4$ taken on an empty stomach is well absorbed, it is very poorly utilized in some fortified foods. A serving of iron-fortified cereal with milk alone may actually provide very little available iron. Practical application of this type of information can help solve problems such as the mysterious occurrence of iron-deficiency anemia in some children on "well-balanced" school meal programs. As long as an iron-fortified cereal was served at 8:00 and orange juice at 10:00, there was a continued incidence of iron deficiency. By moving the orange juice to 8:00 with the cereal, the problem virtually disappeared.

CLIN. CORR. 26.8 IRON-DEFICIENCY ANEMIA

While iron-deficiency anemia is one of the most common nutritional deficiencies in this country, there is some confusion as to just how common it is. The problem is that iron depletion occurs very gradually over a period of time. Iron deficiency occurs in at least three distinct stages. (1) In the first stage, iron stores (usually as ferritin and hemosiderin) are depleted. (2) Once the iron stores have been depleted, serum iron levels fall and the total iron binding capacity (TIBC) of transferrin increases. (3) In the final stage, hemoglobin levels fall and anemia becomes evident. Mean corpuscular hemoglobin concentration (MCHC) is the best indicator of anemia, while % saturation of transferrin (which can be calculated from transferrin and TIBC determinations) is the best measure of simple iron deficiency. Obviously, the % saturation of transferrin is a more sensitive indicator of iron deficiency than MCHC. However, both determinations have sometimes been used interchangeably to estimate the incidence of iron-deficiency anemia.

That leads us to the second question. Just how serious are the symptoms associated with iron-deficiency anemia? In most cases the symptoms are mild enough that the anemia is seldom the reason that the patient comes to see the physician. The most common symptoms are mild fatigue, weakness, anorexia, and pica. In most cases, these symptoms seem to be associated with depleted tissue stores rather than the anemia itself. As the iron-deficiency anemia becomes more severe, the fingernails become thin and flat with a characteristic spoon-shaped appearance (koilonychia). If these were the only pos-

thyroid hormones triiodothyronine and thyroxine. These hormones play a major role in regulating the basal metabolic rate of the adult and the growth and development of the child. Saltwater fish are the best natural food sources of iodine and in earlier years population groups living in inland areas suffered from the endemic deficiency disease goiter. The most characteristic symptom of goiter is the enlargement of the thyroid gland to the point where a large nodule is visible on the neck. Since iodine has been routinely added to table salt, goiter has become relatively rare in this country. However, in some inland areas, mild forms of goiter still may be seen in up to 5% of the population. It is a possible deficiency for individuals not using iodinized salt provided that they have no other dietary sources of iodine.

Zinc

Zinc absorption appears to be dependent on a transport protein, metallothionein. Over 20 zinc metalloenzymes have been described to date, including both RNA and DNA polymerases. Zinc deficiencies in children are usually marked by poor growth and impairment of sexual development. In both children and adults zinc deficiencies result in poor wound healing. Zinc is also present in gustin, a salivary polypeptide that appears to be necessary for normal development of taste buds. Thus zinc deficiencies also lead to decreased taste acuity.

The few dietary surveys that have been carried out in this country have indicated that zinc intake may be marginal for some individuals. However, few symptoms of zinc deficiency other than decreased taste acuity can be demonstrated in those individuals. Severe zinc deficiencies are seen primarily in alcoholics (especially if they have developed cirrhosis), patients with chronic renal disease or severe malabsorption diseases and occasionally in patients on long-term total parenteral nutrition (TPN). The most characteristic early symptom of zinc deficient patients on TPN is dermatitis. Zinc is occasionally used therapeutically to promote wound healing and may be of some use in treating gastric ulcers.

Copper

Copper absorption may also be dependent on the protein metallothionein, since excess intake of either copper or zinc interferes with the absorption of the other. Copper is contained in a number of important metalloenzymes, including cytochrome c oxidase, dopamine β-hydroxylase, superoxide dismutase, and lysyl oxidase. The lysyl oxidase is necessary for the conversion of certain lysine resi-

dues in collagen and elastin to allysine, which is needed for cross-linking. Some of the symptoms of copper deficiency in man are leukopenia, demineralization of bones, anemia, fragility of large arteries, and demyelination of neural tissue. The anemia appears to be due to a defect in iron metabolism. The copper-containing enzyme ferroxidase is necessary for conversion of iron from the Fe^{2+} state (in which form it is absorbed) to the Fe^{3+} state (in which form it can bind to the plasma protein transferrin). The bone demineralization and blood vessel fragility can be directly traced to defects in collagen and elastin formation. The causes of the other symptoms are not known.

Copper balance studies carried out with human volunteers seem to indicate a minimum requirement of 1.5 to 2.0 mg/day. Thus the RDA has been set at 2–3 mg/day. Most dietary surveys find that the average American diet provides only 1 mg at ≦2000 cal/day. At present, this remains a puzzling problem. No symptoms of copper deficiency have been identified in the general public. It is not known whether there exist widespread marginal copper deficiencies, or whether the copper balance studies are inaccurate. Recognizable symptoms of copper deficiency are usually seen only in two relatively rare hereditary diseases, Menke's syndrome and Wilson's disease. Menke's syndrome is associated with a defect in copper transport across cell membranes. Wilson's disease is associated with abnormal accumulation of copper in the tissues and can be treated with the naturally occurring copper chelating agent penicillamine.

Chromium

Chromium probably functions in the body primarily as a component of glucose tolerance factor (GTF), a naturally occurring substance that appears to be a coordination complex between chromium, nicotinic acid, and the amino acids glycine, glutamate, and cysteine. GTF potentiates the effects of insulin, presumably by facilitating its binding to cell receptor sites. The chief symptom of chromium deficiency is impaired glucose tolerance, a result of the decreased insulin effectiveness.

The frequency of occurrence of chromium deficiency is virtually unknown at present. The RDA for chromium has been set at 50–200 μg for a normal adult. The best current estimate is that the average consumption of chromium is around 60 μg/day in the United States. Unfortunately, the range of intakes is very wide (5–100 μg) even for individuals otherwise consuming balanced diets. Those most likely to have marginal or low intakes of chromium are individuals on low caloric intakes or consuming large amounts of processed foods. Some concern has been voiced that many Americans may be marginally deficient in chromium. However, it is difficult to assess the

sible symptoms of iron-deficiency anemia, a major public health effort to prevent iron deficiency would not appear to be warranted. However, there is the possibility that iron deficiency may lead to an increased susceptibility to infection. When tissue stores of iron are depleted, there is an impairment of cell mediated immunity and phagocytic activity. Unfortunately, it is very difficult to accurately correlate the frequency of infection with the iron status in a human population. However, since that possibility clearly exists and a significant portion of the population evidences iron deficiency by one measure or another, several major public health measures have been undertaken to improve iron availability in the diet. These include iron fortification of flour and, more recently, the Women and Infant Children (WIC) Program.

extent of this problem, if it exists, until better chromium analyses of food become available.

The situation is further confused by individual differences in chromium absorption and utilization. For some individuals, GTF appears to be a hormone-like substance in that they can utilize dietary chromium salts, niacin, and amino acids to synthesize GTF. However, other individuals utilize chromium salts very poorly and appear to need preformed GTF in the diet. Unfortunately, very little is known about the requirements for preformed GTF in the general population or about its distribution in natural foods. While it is clear that most diabetics do not respond significantly to either chromium or GTF, there are well documented cases in which GTF has been useful in treating cases of diabetes.

Selenium

Selenium appears to function primarily in the metalloenzyme glutathione peroxidase, which destroys peroxides in the cytosol. Since the effect of vitamin E on peroxide formation is limited primarily to the membrane, both selenium and vitamin E appear to be necessary for efficient scavenging of peroxides. Selenium is one of the few nutrients not removed by the milling of flour and is usually thought to be present in adequate amounts in the diet. The selenium levels are very low in the soil in certain parts of the country, however, and foods raised in these regions will be low in selenium. Fortunately, this effect is minimized by the current food distribution system, which assures that the foods marketed in any one area are derived from a number of different geographical regions.

Other Trace Minerals

Manganese is a component of pyruvate carboxylase and probably other metalloenzymes as well. Molybdenum is a component of xanthine oxidase. Deficiencies of both of these trace minerals are virtually unknown in man. Fluoride is known to strengthen bones and teeth and is usually added to drinking water.

26.10 THE AMERICAN DIET: FACT AND FALLACY

What do dietary surveys tell us about the adequacy of the American diet? The most comprehensive dietary survey presently available is

the HANES Survey (Health and Nutritional Examination Survey), which obtained 24-h recalls from 28,000 Americans age 1 to 74 from 1971 to 1974. It showed that many Americans may be consuming suboptimal amounts of iron, calcium, vitamin A, and vitamin C. Less extensive dietary surveys have suggested a significant fraction of the population might have inadequate intakes of B_6, folic acid, and certain trace minerals. How are these data to be interpreted? In every instance, biochemical measurements show significantly fewer individuals with marginal nutritional status, and clinical symptoms of these deficiencies are rare indeed. Thus a physician need not be alarmed by these reports of potential dietary deficiencies, but should be aware of them when dealing with patients with increased nutrient requirements.

Much has been said lately about the supposed deterioration of the American diet. How serious a problem is this? Clearly Americans are eating much more processed food than our ancestors. These foods differ from simpler foods in that they have a higher caloric density and a lower nutrient density than the foods they replace. However, these foods are almost uniformly enriched with iron, thiamine, riboflavin, and niacin. In many cases they are even fortified (usually as much for sales promotion as for nutritional reasons) with as many as 11–15 vitamins and minerals. Unfortunately, it is simply not practical to replace all of the nutrients lost, especially the trace minerals. Imitation foods present a special problem in that they are usually incomplete in more subtle ways. For example, the imitation cheese and imitation milkshakes that are widely sold in this country usually do contain the protein and calcium one would expect of the food they replace, but often do not contain the riboflavin, which one should also obtain from these items. Fast food restaurants have also been much maligned in recent years. Some of the criticism has been undeserved, but fast food meals do tend to be high in calories and fat and low in certain vitamins and trace minerals. For example, the standard fast food meal provides over 50% of the calories the average adult needs for the entire day, while providing $<5\%$ of the vitamin A and $<30\%$ of biotin, folic acid, and pantothenic acid. Unfortunately, much of the controversy in recent years has centered around whether these trends are "good" or "bad." This simply obscures the issue at hand. Clearly it is possible to obtain a balanced diet which includes processed, imitation, and fast foods if one compensates by selecting foods for the other meals which are low in caloric density and rich in nutrients. Unfortunately, few nutritionists have taken the initiative in pointing out that such a compensation is both necessary and possible if one wishes to consume a balanced diet.

What then are the important changes in nutrient intake in the

American diet? Clearly the decreased intake of some micronutrients that one might expect has been partially balanced by enrichment and fortification. There has been little or no decrease in intake of iron, thiamine, riboflavin, niacin, ascorbic acid, vitamin A, and vitamin B_{12}. However, those nutrients that are not commonly added to enriched or fortified foods do appear to be decreasing. These include vitamin B_6, folic acid, trace minerals, and fiber. A certain amount of concern perhaps is justifiable over this trend. However, there are instances in which the physician should be particularly aware of the potential shortfalls in nutrient intake. For example, the difference between actual intakes of iron and calcium and probable need for these nutrients in infants and pregnant/lactating mothers is clearly a cause of concern. Likewise marginal intake of folates and B_6 in many young females of childbearing age may increase their risk of folate and B_6 deficiency during pregnancy and B_6 deficiency when using birth control pills. Although deficient calcium intake is clearly not a cause of osteoporosis, there should be some concern that continued inadequate calcium intake in postmenopausal women may accelerate the rate of osteoporosis.

CLIN. CORR. 26.9 NUTRITIONAL CONSIDERATIONS FOR VEGETARIANS

A vegetarian diet poses certain problems in terms of micronutrient intake which need to be recognized in designing a well-balanced diet. Vitamin B_{12} is of special concern, since it is found only in foods of animal origin. B_{12} should be obtained from fortified foods (such as some brands of soybean milk) or in tablet form. However, surprisingly few vegetarians ever develop pernicious anemia, perhaps because an adult who has previously eaten meat will have a 6- to 10-year store of B_{12} in his liver.

26.11 ASSESSMENT OF NUTRITIONAL STATUS IN CLINICAL PRACTICE

Having surveyed the major micronutrients and their biochemical roles, it might seem that the process of evaluating the nutritional status of an individual patient would be an overwhelming task. Perhaps the most useful skill for the clinician will be the recognition of high-risk population groups, that is, the patients most likely to be at nutritional risk. The population groups with the poorest dietary intake tend to be low income groups from inner city areas (especially Spanish-Americans and blacks) and depressed rural areas (such as Appalachia). Their diets are most likely to be deficient in iron, calcium, vitamin A, and vitamin C. The elderly also tend to have poor nutrient intake due to restricted income, loss of appetite, and loss of ability to prepare a wide variety of foods. One should also be aware of the nutrients most consistently below two-thirds of the RDA for teenagers (iron, calcium, vitamin C, and vitamin A) and for adult females (iron and calcium). In interpreting the risk associated with poor dietary intake, the physician should be aware of those population groups with the greatest nutritional needs. For example, infants

and young children have increased needs for iron, calcium, and protein. Iron and calcium are most likely to be a problem for this age group. Pregnant and lactating women have increased needs for protein, vitamins (especially folic acid and B_6) and minerals (especially iron and calcium). Vegetarian diets are most likely to be deficient in calcium, iron, and vitamins D and B_{12} (Clin. Corr. 26.9).

Illness and metabolic stress often cause increased demand or decreased utilization of certain nutrients. For example, diseases leading to fat malabsorption cause a particular problem with absorption of calcium and the fat-soluble vitamins. Other malabsorption diseases can result in deficiencies of many nutrients depending on the particular malabsorption disease. Liver and kidney disease can prevent activation of vitamin D and storage or utilization of many other nutrients including vitamin A, vitamin B_{12}, and folic acid. Severe illness or trauma increase the need for calories, protein, and possibly some micronutrients such as vitamin C and certain B vitamins. Long-term use of many drugs in the treatment of chronic disease states can affect the need for certain micronutrients. Some of these are summarized in Table 26.1.

Table 26.1 Drug-Nutrient Interactions

Drug	*Potential Nutrient Deficiencies*	*Observed Clinical Effect*
Alcohol	Folic acid	Anemia
	Vitamin B_6	Sideroblastic anemia
		Peripheral neuropathy (?)
	Thiamine	Wernicke–Korsakoff Syndrome
Anticonvulsants	Vitamin D	Rickets
	Folic acid	Megaloblastic anemia
	Vitamin K	Neonatal hemorrhaging
Corticosteroids	Vitamin D and calcium	Accelerated bone loss
	Zinc	Decreased wound healing
	Potassium	Muscle weakness
	Vitamin B_6	Decreased glucose tolerance (?)
	Vitamin C	Decreased wound healing (?)
		Accelerated bone loss (?)
Diuretics	Potassium	Muscle weakness
	Zinc	Decreased wound healing
	Calcium, magnesium	
Oral contraceptives and estrogens	Vitamin B_6	Mental depression
		Abnormal glucose tolerance
	Folic acid	Megaloblastic changes
	Vitamins C and B_{12}	

Iron is another problem. The best vegetable sources of iron are dried beans, dried fruits, whole grain or enriched cereals, and green leafy vegetables. Vegetarian diets can provide adequate amounts of iron provided that these foods are regularly selected and consumed with vitamin C-rich foods to promote iron absorption (Clin. Corr. 26.7). However, iron supplementation is usually recommended for children and menstruating females.

When milk and dairy products are absent from the diet, certain other problems must be considered as well. Normally, dietary vitamin D is obtained primarily from fortified milk. While some butters and margarines are fortified with vitamin D, they are seldom consumed in sufficient quantities to supply significant amounts of vitamin D. Although adults can usually obtain sufficient vitamin D from exposure to sunlight, dietary sources are usually necessary during periods of growth and for adults with little exposure to sunlight. Vegetarians may need to obtain their vitamin D from fortified foods such as cereals, certain soybean milks, or in tablet form. Riboflavin is found in a number of vegetable sources such as green leafy vegetables, enriched breads, and wheat germ. However, since none of these sources supply more than 10% of the RDA in normal serving sizes, fortified cereals or vitamin supplements may become an important source of this nutrient. The important sources of calcium for vegetarians include soybeans, soybean milk, almonds, and green leafy vegetables. Those green leafy vegetables without oxalic acid (mustard, turnip, and dandelion greens, collards, kale, romaine, and loose leaf lettuce) are particularly good sources of calcium. However, none of these sources is equivalent to cow's milk in calcium content, so calcium supplements are usually recommended during periods of rapid growth.

Who then is at nutritional risk? Obviously, this depends on many factors. Nutritional counseling will be an important part of the treatment for infants, young children, and pregnant/lactating females. A brief analysis of a dietary history and further nutritional counseling will also be important when dealing with certain other high-risk patients.

BIBLIOGRAPHY

Recommended Dietary Allowances

Food and Nutrition Board of the National Academy of Sciences. *Recommended dietary allowances,* 9th ed. Washington, D.C.: National Academy of Sciences, 1980.

Micronutrients—General Information

The Nutrition Foundation. *Present knowledge in nutrition,* 4th ed. New York, 1976.

Vitale, J. J. *Vitamins.* Kalamazoo, Mich.: Upjohn Co., 1976.

Vitamin E

Howitt, M. K. Vitamin E: a reexamination. *Am. J. Clin. Nutr.* 29:569, 1976.

Vitamin C

Anderson, T. W. Large scale trials of vitamin C. *Ann. N.Y. Acad. Sci.* 258:498, 1975.

Calcium

Heaney, R. P., Recker, R. R., and Saville, P. D. Calcium balance studies and calcium requirements in middle-aged women. *Am. J. Clin. Nutr.* 30:1603, 1977.

Vitamin B_6 and Oral Contraceptives

Kirksey, A., Keaton, K., Abernathy, R. P., and Greger, J. L. Vitamin B_6 nutritional status of a group of female adolescents. *Am. J. Clin. Nutr.* 31:946, 1978.

Bossé, T. R., and Donald, E. A. The vitamin B_6 requirement in oral contraceptive uses. *Am. J. Clin. Nutr.* 32:1015, 1979.

Dietary Iron

Manson, E. R., Hallberg, L., Layrisse, M., Hegsted, D. M., Cook, J. D., Mertz, W., and Finch, C. A. Estimation of dietary iron. *Am. J. Clin. Nutr.* 31:134, 1978.

Nutrition and Alcoholism

Shaw, S., and Lieber, C. S. Alcoholism, in H. Schneider, C. Anderson, and D. Coursin (eds.), *Nutritional support of medical practice.* New York: Harper & Row, 1977, p. 202.

Drug-Nutrient Interaction

Theuer, R. C., and Vitale, J. J. Drug and nutrient interactions, in H. Schneider, C. Anderson, and D. Coursin (eds.), *Nutritional support of medical practice.* New York: Harper & Row, 1977, p. 297.

Dietary Surveys

U.S. Department of Health, Education, and Welfare. *Highlights of the ten-state nutrition survey,* DHEW Publication No. 72-8134. Washington, D.C.: U.S. Government Printing Office, 1972.

U.S. Department of Health, Education, and Welfare. *Preliminary findings of the first health and nutrition examination survey, 1971–1973,* DHEW Publication No. (HRA) 74-1219-1. Washington, D.C.: U.S. Government Printing Office, 1974.

Index

Page numbers followed by CC refer to material under Clinical Correlation sections.

This Wiley International Edition is part of a continuing program of paperbound textbooks especially designed for students and professional people overseas. It is an unabridged reprinting of the original hardbound edition, which is also available from your bookseller.

Wiley International Editions include titles in the fields of:

- Agricultural Engineering & Agriculture
- Anthropology
- Biochemistry
- Biology
- Business Administration
- Chemistry
- Civil Engineering
- Chemical Engineering
- Computers & Data Processing
- Earth Sciences
- Economics
- Education
- Electrical Engineering
- Engineering Mechanics
- Geography
- Home Economics
- Industrial Engineering
- Mathematics
- Materials Engineering
- Mechanical Engineering
- Medicine
- Physics
- Physical Chemistry
- Polymer Science & Technology
- Probability & Statistics
- Psychology
- Sociology
- Vocational-Technical

JOHN WILEY & SONS, Inc.
605 THIRD AVENUE
NEW YORK, NEW YORK 10016

ISBN 0-471-87286-5