

## Preface

The authors and publishers are pleased to present the twenty-ninth edition of *Harper's Illustrated Biochemistry*. The first edition of this text, entitled *Harper's Biochemistry*, was published in 1939 under the sole authorship of Dr Harold Harper, University of California, San Francisco. Subsequently, various authors have contributed to the text.

## Cover Illustration for the Twenty-Ninth Edition

The cover illustration for the 29th edition commemorates Elizabeth H. Blackburn, Carol W. Greider, and Jack W. Szostak, who shared the 2009 Nobel Prize in Physiology or Medicine for their seminal work on telomeres and the enzyme telomerase. Telomeres comprise up to 200 copies of a repeating DNA sequence called a G-quadruplex, a structure named for the unique cyclic arrangement of four sets of four guanine bases hydrogen-bonded in head-to-tail fashion that stabilize this structure. In the illustration, the phosphodiester backbone of the DNA is represented by a ribbon and the guanine bases by filled hexagons fused to filled pentagons. The spectral color gradation from purple to red facilitates tracing the progression of the polynucleotide chain. The four sets of cyclic tetra-guanine units can be seen in center stacked from top to bottom and tilted roughly 45° from left to right (Adapted from Protein Data Bank ID no. 2KKA).

As a consequence of the unidirectional nature of DNA replication, each time a chromosome is replicated, the number of G-quadruplex units is reduced. When the supply of telomere units is completely exhausted, replication ceases and the cell transitions to a senescent state. Scientists speculate that the telomere serves as a countdown clock that limits the number of times a somatic cell can divide, and hence its lifespan.

## Changes in the Twenty-Ninth Edition

Consistent with our goal of providing students with a text that describes and illustrates biochemistry in a medically relevant, up-to-date, comprehensive, and yet relatively concise manner, in addition to updating every chapter, significant new material appears in this edition.

Each chapter now begins with a brief statement of its objectives followed by a brief account of its biomedical importance. A major addition is the inclusion of over 250 multiple-choice exam questions with answers given in an answer bank.

## Major Additional Changes Include Three Entirely New Chapters:

"Biochemistry of Aging"  
"Biochemistry of Cancer"  
"Clinical Chemistry"

## Additional Significant Changes Include:

- Inclusion of aspects of epidemiology in the chapter on "Bioinformatics and Computational Biology".
- New figures that illustrate key approaches for identifying possible active sites, ligand-binding sites, and other interaction sites (Section I), and various aspects of metabolism (Section II).
- New tables that summarize aspects of metabolic diseases, including those of purine, pyrimidine, and amino acid metabolism (Section III).
- Expanded discussion of non-coding RNAs, DNA damage repair and human diseases, epigenetic factors that control eukaryotic gene expression, the activities of miRNAs, and powerful new assays to monitor and characterize transcription genome-wide (Section IV).
- New tables that address vitamin and mineral requirements and a greatly expanded discussion of iron metabolism in health and disease (Section VI).

## Organization of the Book

Following two introductory chapters, the text is divided into six main sections. All sections and chapters emphasize the medical relevance of biochemistry.

**Section I** addresses the structures and functions of proteins and enzymes. This section also contains a chapter on Bioinformatics and Computational Biology, reflecting the increasing importance of these topics in modern biochemistry, biology, and medicine.

**Section II** explains how various cellular reactions either utilize or release energy, and traces the pathways by which carbohydrates and lipids are synthesized and degraded. Also described are the many functions of these molecules.

**Section III** deals with the amino acids, their metabolic fates, certain features of protein catabolism, and the biochemistry of the porphyrins and bile pigments.

**Section IV** describes the structure and function of nucleotides and nucleic acids, DNA replication and repair, RNA synthesis and modification, protein synthesis, the principles of recombinant DNA technology, and new understanding of how gene expression is regulated.

**Section V** deals with aspects of extracellular and intracellular communication. Topics include membrane structure and function, the molecular bases of the actions of hormones, and the field of signal transduction.

**Section VI** includes fifteen special topics: nutrition, digestion, and absorption; vitamins and minerals; free radicals and antioxidants; intracellular trafficking and sorting of proteins; glycoproteins; the extracellular matrix; muscle and the cytoskeleton; plasma proteins and immunoglobulins; hemostasis and thrombosis; red and white blood cells; the metabolism of xenobiotics; the biochemistry of aging; the biochemistry of cancer; clinical chemistry; and sixteen biochemically oriented case histories. The latter chapter concludes with a brief epilog indicating some major

challenges for medicine for which biochemistry and related disciplines will play important roles in finding solutions.

**Appendix** lists useful web sites and biochemical journals and others with significant biochemical content.

### Acknowledgments

The authors thank Michael Weitz for his role in the planning of this edition, and Brian Kearns for his key role in getting this edition ready for publication. We also thank Mala Arora and her colleagues at Thomson Digital for their efforts in editing, typesetting, and artwork, and Calvin "Nic" Steussy of Purdue University for his assistance in generating the cover illustration.

Suggestions from students and colleagues around the world have been most helpful in the formulation of this edition. We look forward to receiving similar input in the future.

Rob Murray acknowledges with thanks Joe Varghese and Molly Jacob as co-authors of Chapters 50, 55, and 56, Fred Keeley for his many contributions to Chapter 48, Peter Gross for co-authorship of Chapters 51 and 57, and Margaret Rand for co-authorship of Chapter 51. Special thanks are extended to Reinhart Reithmeier, Alan Volchuk, and David Williams for reviewing and making invaluable suggestions for the revision of Chapters 40 and 46.

Robert K. Murray

David A. Bender

Kathleen M. Botham

Peter J. Kennelly

Victor W. Rodwell

P. Anthony Weil

## **Copyright Information**

### **Harper's Illustrated Biochemistry, Twenty-Ninth Edition**

Copyright © 2012 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in China. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

ISBN 978-0-07-176576-3

MHID 0-07-176576-X

ISSN 1043-9811

### **Notice**

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The authors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possibility of human error or changes in medical sciences, neither the authors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they disclaim all responsibility for any errors or omissions or for the results obtained from use of the information contained in this work. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this work is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

## **Contributors**

### **Editors**

#### **Robert K. Murray, MD, PhD**

Emeritus Professor of Biochemistry  
University of Toronto  
Toronto, Ontario

#### **David A. Bender, PhD**

Professor (Emeritus) of Nutritional Biochemistry  
University College London  
London, United Kingdom

#### **Kathleen M. Botham, PhD, DSc**

Professor of Biochemistry  
Department of Veterinary Basic Sciences  
Royal Veterinary College  
University of London  
London, United Kingdom

#### **Peter J. Kennelly, PhD**

Professor and Head  
Department of Biochemistry  
Virginia Polytechnic Institute and State University  
Blacksburg, Virginia

#### **Victor W. Rodwell, PhD**

Professor (Emeritus) of Biochemistry  
Purdue University  
West Lafayette, Indiana

#### **P. Anthony Weil, PhD**

Professor of Molecular Physiology and Biophysics  
Vanderbilt University School of Medicine  
Nashville, Tennessee

### **Co-Authors**

#### **Daryl K. Granner, MD**

Professor (Emeritus) of Molecular Physiology and Biophysics and Medicine  
Vanderbilt University  
Nashville, Tennessee

#### **Peter L. Gross, MD, MSc, FRCP(C)**

Associate Professor  
Department of Medicine  
McMaster University  
Hamilton, Ontario

#### **Molly Jacob, MB BS, MD, PhD**

Professor and Chair  
Department of Biochemistry  
Christian Medical College  
Vellore, Tamil Nadu

#### **Frederick W. Keeley, PhD**

Associate Director and Senior Scientist  
Research Institute, Hospital for Sick Children, Toronto, and Professor of Biochemistry  
Department of Biochemistry  
University of Toronto  
Toronto, Ontario

#### **Peter A. Mayes, PhD, DSc**

Emeritus Professor of Veterinary Biochemistry  
Royal Veterinary College  
University of London  
London, United Kingdom

#### **Margaret L. Rand, PhD**

Associate Senior Scientist  
Hospital for Sick Children  
Toronto and Professor  
Departments of Laboratory  
Medicine & Pathobiology and Biochemistry  
University of Toronto  
Toronto, Ontario

**Joe Varghese, MB BS, MD**  
Assistant Professor  
Department of Biochemistry  
Christian Medical College  
Vellore, Tamil Nadu

## OBJECTIVES

After studying this chapter, you should be able to:

- Explain what biochemistry is about and appreciate its central role in the life sciences.
- Understand the relationship of biochemistry to health and disease and to medicine.
- Appreciate how the Human Genome Project has given rise to, or stimulated interest in numerous disciplines that are already illuminating many aspects of biology and medicine.

## INTRODUCTION

Biochemistry can be defined as *the science of the chemical basis of life* (Gk *bios* "life"). The **cell** is the structural unit of living systems. Thus, biochemistry can also be described as *the study of the chemical constituents of living cells and of the reactions and processes they undergo*. By this definition, biochemistry encompasses large areas of **cell biology**, **molecular biology**, and **molecular genetics**.

### The Aim of Biochemistry Is to Describe and Explain, in Molecular Terms, All Chemical Processes of Living Cells

The **major objective** of biochemistry is **the complete understanding, at the molecular level, of all of the chemical processes associated with living cells**. To achieve this objective, biochemists have sought to isolate the numerous molecules found in cells, determine their structures, and analyze how they function. Many techniques have been used for these purposes; some of them are summarized in **Table 1-1**.

**Table 1-1 The Principal Methods and Preparations Used in Biochemical Laboratories**

#### Methods for Separating and Purifying Biomolecules<sup>1</sup>

Salt fractionation (eg, precipitation of proteins with ammonium sulfate)  
 Chromatography: Paper, ion exchange, affinity, thin-layer, gas-liquid, high-pressure liquid, gel filtration  
 Electrophoresis: Paper, high-voltage, agarose, cellulose acetate, starch gel, polyacrylamide gel, SDS-polyacrylamide gel  
 Ultracentrifugation

#### Methods for Determining Biomolecular Structures

Elemental analysis  
 UV, visible, infrared, and NMR spectroscopy  
 Use of acid or alkaline hydrolysis to degrade the biomolecule under study into its basic constituents  
 Use of a battery of enzymes of known specificity to degrade the biomolecule under study (eg, proteases, nucleases, glycosidases)  
 Mass spectrometry  
 Specific sequencing methods (eg, for proteins and nucleic acids)  
 X-ray crystallography

#### Preparations for Studying Biochemical Processes

Whole animal (includes transgenic animals and animals with gene knockouts)  
 Isolated perfused organ  
 Tissue slice  
 Whole cells  
 Homogenate  
 Isolated cell organelles  
 Subfractionation of organelles  
 Purified metabolites and enzymes  
 Isolated genes (including polymerase chain reaction and site-directed mutagenesis)

<sup>1</sup>Most of these methods are suitable for analyzing the components present in cell homogenates and other biochemical preparations. The sequential use of several techniques will generally permit purification of most biomolecules. The reader is referred to texts on methods of biochemical research for details.

Other objectives of biochemistry include helping to **understand the origins of life on Earth** and to integrate biochemical knowledge into efforts to **Maintain health** and to **understand diseases and treat them effectively**.

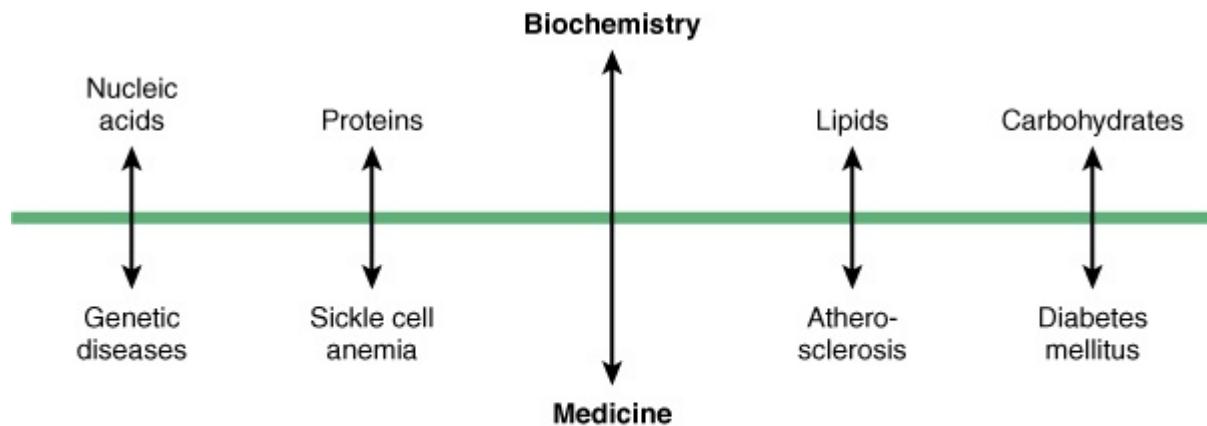
## A Knowledge of Biochemistry Is Essential to All Life Sciences

The biochemistry of the nucleic acids lies at the heart of **genetics**; in turn, the use of genetic approaches has been critical for elucidating many areas of biochemistry. **Cell biology** is very closely allied to biochemistry. **Physiology**, the study of body function, overlaps with biochemistry almost completely. **Immunology** employs numerous biochemical techniques, and many immunologic approaches have found wide use by biochemists. **Pharmacology** and **pharmacy** rest on a sound knowledge of biochemistry and physiology; in particular, most drugs are metabolized by enzyme-catalyzed reactions. Poisons act on biochemical reactions or processes; this is the subject matter of **toxicology**. Biochemical approaches are being used increasingly to study basic aspects of **pathology** (the study of disease), such as inflammation, cell injury, and cancer. Many workers in **microbiology**, **zoology**, and **botany** employ biochemical approaches almost exclusively. These relationships are not surprising, because life as we know it depends on biochemical reactions and processes. In fact, the old barriers among the life sciences are breaking down, and biochemistry is increasingly becoming their **common language**.

## A Reciprocal Relationship between Biochemistry & Medicine Has Stimulated Mutual Advances

The two major concerns for workers in the health sciences—and particularly physicians—are the understanding and maintenance of **health** and the understanding and effective treatment of **diseases**. Biochemistry impacts enormously on both of these fundamental concerns of medicine. In fact, the interrelationship of biochemistry and medicine is a wide, two-way street. Biochemical studies have illuminated many aspects of health and disease, and conversely, the study of various aspects of health and disease has opened up new areas of biochemistry. Some examples of this two-way street are shown in **Figure 1-1**. For instance, knowledge of protein structure and function was necessary to elucidate the single biochemical difference between **normal hemoglobin** and **sickle cell hemoglobin**. On the other hand, analysis of sickle cell hemoglobin has contributed significantly to our understanding of the structure and function of both normal hemoglobin and other proteins. Analogous examples of reciprocal benefit between biochemistry and medicine could be cited for the other paired items shown in **Figure 1-1**. Another example is the pioneering work of Archibald Garrod, a physician in England during the early 1900s. He studied patients with a number of relatively rare disorders (alkaptonuria, albinism, cystinuria, and pentosuria; these are described in later chapters) and established that these conditions were genetically determined. Garrod designated these conditions as **inborn errors of metabolism**. His insights provided a major foundation for the development of the field of human biochemical genetics. More recent efforts to understand the basis of the genetic disease known as **familial hypercholesterolemia**, which results in severe atherosclerosis at an early age, have led to dramatic progress in understanding of cell receptors and of mechanisms of uptake of cholesterol into cells. Studies of **oncogenes** and **tumor suppressor genes** in cancer cells have directed attention to the molecular mechanisms involved in the control of normal cell growth. These and many other examples emphasize how the study of disease can open up areas of cell function for basic biochemical research.

**Figure 1-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Examples of the two-way street connecting biochemistry and medicine.** Knowledge of the biochemical molecules shown in the top part of the diagram has clarified our understanding of the diseases shown on the bottom half—and conversely, analyses of the diseases shown below have cast light on many areas of biochemistry. Note that sickle cell anemia is a genetic disease and that both atherosclerosis and diabetes mellitus have genetic components.

The relationship between medicine and biochemistry has important implications for the former. As long as medical treatment is firmly grounded in the knowledge of biochemistry and other basic sciences, the practice of medicine will have a **rational basis** that can be adapted to accommodate new knowledge. This contrasts with unorthodox health cults and at least some "alternative medicine" practices that are often founded on little more than myth and wishful thinking and generally lack any intellectual basis.

Biochemistry is one important area of science. The many ways in which **science is important for physicians** (and equally so for other workers in health care or biology, whether concerned with humans or animals) have been well stated in an article by Cooke (2010). They include (i) offering a foundational understanding on which one's practice should be built, (ii) stimulating curiosity and creating the scientific habits that are essential for continual learning throughout one's career, (iii) showing how our present knowledge has been acquired, and (iv) emphasizing the immensity of what is as yet unknown. Of course, it is vital that the application of science to helping a patient must be practised with humanity and the highest ethical standards.

## NORMAL BIOCHEMICAL PROCESSES ARE THE BASIS OF HEALTH

The World Health Organization (WHO) defines **health** as a state of "complete physical, mental, and social well-being and not merely the absence of disease and infirmity." From a strictly biochemical viewpoint, health may be considered that situation in which all of the many thousands of intra- and extracellular reactions that occur in the body are proceeding at rates commensurate with the organism's maximal survival in the physiologic state. However, this is an extremely reductionist view, and it should be apparent that caring for the health of patients requires not only a wide knowledge of **biologic principles** but also of **psychologic** and **social** principles.

### Biochemical Research Has Impact on Nutrition & Preventive Medicine

One major prerequisite for the maintenance of health is that there be optimal dietary intake of a number of chemicals; the chief of these are **vitamins**, certain **amino acids**, certain **fatty acids**, various **minerals**, and **water**. Because much of the subject matter of both **biochemistry and nutrition** is concerned with the study of various aspects of these chemicals, there is a close relationship between these two sciences. Moreover, more emphasis is being placed on systematic attempts to maintain health and forestall disease, that is, on **preventive medicine**. Thus, nutritional approaches to—for example—the prevention of atherosclerosis and cancer are receiving increased emphasis. Understanding nutrition depends to a great extent on knowledge of biochemistry.

### Most & Perhaps All Diseases Have a Biochemical Basis

We believe that most if not all diseases are manifestations of abnormalities of molecules, chemical reactions, or biochemical processes. The **major factors responsible for causing diseases** in animals and humans are listed in **Table 1–2**. All of them affect one or more critical chemical reactions or molecules in the body. Numerous examples of the biochemical bases of diseases will be encountered in this text. In most of these conditions, biochemical studies contribute to both the diagnosis and treatment. Some **major uses of biochemical investigations and of laboratory tests in relation to diseases** are summarized in **Table 56–1**. Chapter 56 describes many aspects of the field of **clinical biochemistry**, which is mainly concerned with the use of biochemical tests to assist in the diagnosis of disease and also in the overall management of patients with various disorders. Chapter 57 further helps to illustrate the relationship of biochemistry to disease by discussing in some detail biochemical aspects of 16 different medical cases.

**Table 1–2 The Major Causes of Diseases<sup>1</sup>**

1. **Physical agents:** Mechanical trauma, extremes of temperature, sudden changes in atmospheric pressure, radiation, electric shock.
2. **Chemical agents, including drugs:** Certain toxic compounds, therapeutic drugs, etc.
3. **Biologic agents:** Viruses, bacteria, fungi, higher forms of parasites.
4. **Oxygen lack:** Loss of blood supply, depletion of the oxygen-carrying capacity of the blood, poisoning of the oxidative enzymes.
5. **Genetic disorders:** Congenital, molecular.
6. **Immunologic reactions:** Anaphylaxis, autoimmune disease.
7. **Nutritional imbalances:** Deficiencies, excesses.
8. **Endocrine imbalances:** Hormonal deficiencies, excesses.

<sup>1</sup>**Note:** All of the causes listed act by influencing the various biochemical mechanisms in the cell or in the body.

**Source:** Adapted, with permission, from Robbins SL, Cotran RS, Kumar V: *The Pathologic Basis of Disease*, 3rd ed. Saunders, 1984. Copyright © 1984 Elsevier Inc. with permission from Elsevier.

Some of the **major challenges that medicine and related health sciences** face are also outlined very briefly at the end of Chapter 57. In addressing these challenges, biochemical studies are already and will continue to be interwoven with studies in various other disciplines, such as genetics, cell biology, immunology, nutrition, pathology, and pharmacology. Many biochemists are vitally interested in contributing to solutions to key issues such as how can the survival of mankind be assured, and also in educating the public to support the use of the scientific method in solving major problems (eg, environmental and others) that confront us.

### Impact of the Human Genome Project (HGP) on Biochemistry, Biology, & Medicine

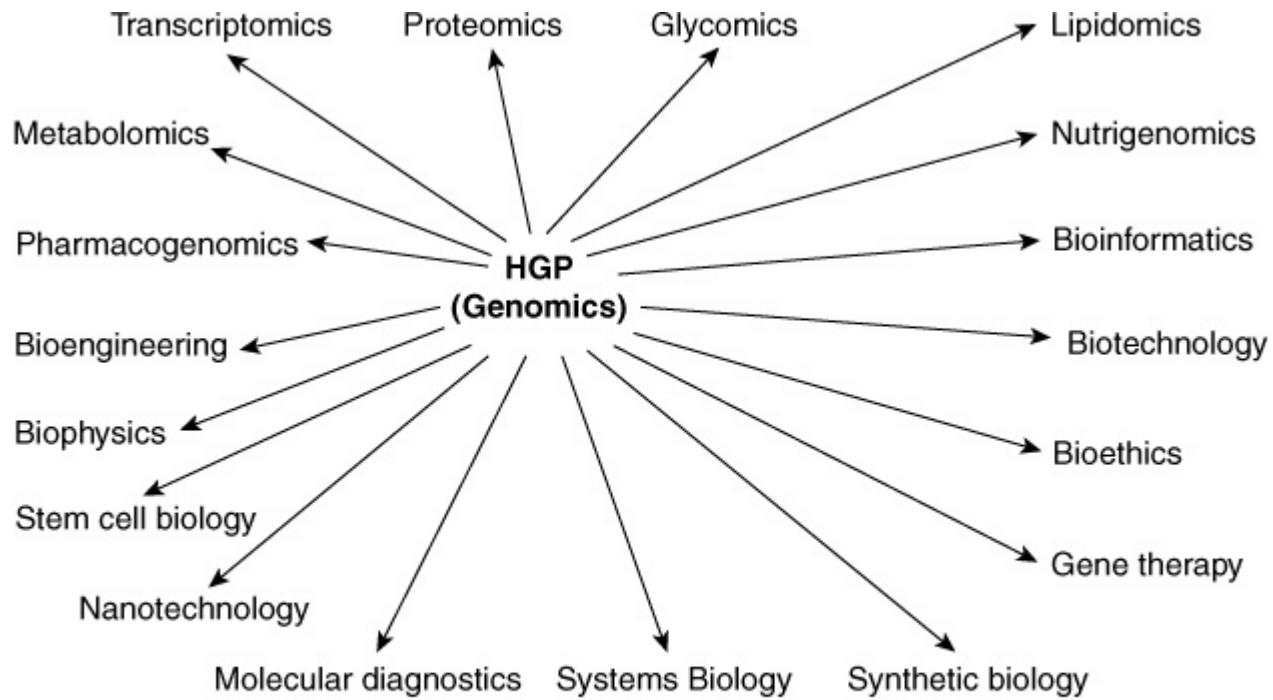
Remarkable progress was made in the late 1990s in sequencing the human genome by the HGP. This culminated in July 2000, when leaders of the two groups involved in this effort (the International Human Genome Sequencing Consortium and Celera Genomics, a private company) announced that over 90% of the genome had been sequenced. Draft versions of the sequence were published in early 2001. With the exception of a few gaps, the sequence of the entire human genome was completed in 2003, 50 years after the description of the double-helical nature of DNA by Watson and Crick.

The **implications of the HGP for biochemistry**, all of **biology**, and for **medicine and related health sciences** are tremendous, and only a few points are mentioned here. It is **now possible to isolate any gene and usually determine its structure and function** (eg, by sequencing and knockout experiments). Many **previously unknown genes** have been revealed; their products have already been established, or are under study. New light has been thrown on **human evolution**, and procedures for **tracking disease genes** have been greatly refined. Reference to the HGP will be made in various chapters of this text.

As the ramifications of the HGP increase, it is vital for readers to understand **the major contributions to understanding human health and disease** that have been made, and are being made, by **studies of the genomes of model organisms**, particularly *Drosophila melanogaster* (the fruit fly) and *Caenorhabditis elegans* (the round worm). This has been clearly stated by Bruce Alberts (2010) in reflecting on the recent impressive progress made in deciphering the genomes of these two organisms. Because these organisms can be experimentally manipulated and have short generation times, relatively rapid progress can be made in understanding the normal functions of their genes and also how abnormalities of their genes can cause disease. Hopefully these advances can be translated into approaches that help humans. According to Alberts, "As incredible as it seems, future research on flies and worms will quite often provide the shortest and most efficient path to curing human diseases." This applies to disorders as different as cancer and Alzheimer disease.

**Figure 1–2** shows **areas of great current interest** that have developed either directly as a result of the progress made in the HGP, or have been spurred on by it. As an outgrowth of the HGP, many so-called **-omics** fields have sprung up, involving comprehensive studies of the structures and functions of the molecules with which each is concerned. Definitions of the fields listed below are given in the Glossary of this chapter. The products of genes (RNA molecules and proteins) are being studied using the technics of **transcriptomics** and **proteomics**. One spectacular example of the speed of progress in transcriptomics is the explosion of knowledge about small RNA molecules as regulators of gene activity. Other -omics fields include **glycomics**, **lipidomics**, **metabolomics**, **nutrigenomics**, and **pharmacogenomics**. To keep pace with the amount of information being generated, **bioinformatics** has received much attention. Other related fields to which the impetus from the HGP has carried over are **biotechnology**, **bioengineering**, **biophysics**, and **bioethics**. **Nanotechnology** is an active area, which, for example, may provide novel methods of diagnosis and treatment for cancer and other disorders. **Stem cell biology** is at the center of much current research. **Gene therapy** has yet to deliver the promise that it offers, but it seems probable that will occur sooner or later. Many new **molecular diagnostic tests** have developed in areas such as genetic, microbiologic, and immunologic testing and diagnosis. **Systems biology** is also burgeoning. **Synthetic biology** is perhaps the most intriguing of all. This has the potential for creating living organisms (eg, initially small bacteria) from genetic material in vitro. These could perhaps be designed to carry out specific tasks (eg, to mop up petroleum spills). As in the case of stem cells, this area will attract much attention from bioethicists and others. Many of the above topics are referred to later in this text.

**Figure 1–2**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**The Human Genome Project (HGP) has influenced many disciplines and areas of research.** Biochemistry itself is not shown in this figure, because it was underway long before the HGP commenced. However, a number of the disciplines shown (eg, bioinformatics, genomics, glycomics, lipidomics, metabolomics, molecular diagnostics, proteomics, and transcriptomics) are very active areas of research by biochemists.

All of the above have made the present time a very exciting one for studying or to be directly involved in biology and medicine. The outcomes of research in the various areas mentioned above will impact tremendously on the future of biology, medicine, and the health sciences.

## SUMMARY

- Biochemistry is the science concerned with studying the various molecules that occur in living cells and organisms and with their chemical reactions. Because life depends on biochemical reactions, biochemistry has become the basic language of all biologic sciences.
- Biochemistry is concerned with the entire spectrum of life forms, from relatively simple viruses and bacteria to complex human beings.
- Biochemistry and medicine and other health care disciplines are intimately related. Health in all species depends on a harmonious balance of biochemical reactions occurring in the body, and disease reflects abnormalities in biomolecules, biochemical reactions, or biochemical processes.
- Advances in biochemical knowledge have illuminated many areas of medicine. Conversely, the study of diseases has often revealed previously unsuspected aspects of biochemistry. Biochemical approaches are often fundamental in illuminating the causes of diseases and in designing appropriate therapies.
- The judicious use of various biochemical laboratory tests is an integral component of diagnosis and monitoring of treatment.
- A sound knowledge of biochemistry and of other related basic disciplines is essential for the rational practice of medicine and related health sciences.
- Results of the HGP and of research in related areas will have a profound influence on the future of biology, medicine, and other health sciences. The importance of genomic research on model organisms such as *D melanogaster* and *C elegans* for understanding human diseases is emphasized.

## REFERENCES

- Alberts B: Model organisms and human health. *Science* 2010;330:1724.
- Alberts B: Lessons from genomics. *Science* 2011;331:511. (In this issue of *Science* and succeeding issues in February 2011 various scientists comment on the significance of the tenth anniversary of the publications of the sequencing of the human genome).
- Cammack R, Attwood T, Campbell P, et al (editors): *Oxford Dictionary of Biochemistry and Molecular Biology*. 2nd ed. Oxford University Press. 2006.
- Cooke M. Science for physicians. *Science* 2010;329;1573.
- Feero WG, Guttmacher AE, Collins FS: Genomic medicine—an updated primer. *N Eng J Med* 2010;362:2001.
- Fruton JS: *Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology*. Yale University Press, 1999. (Provides the historical background for much of today's biochemical research.)
- Garrod AE: Inborn errors of metabolism. (*Croonian Lectures*.) *Lancet* 1908;2:1:73,142,214.
- Gibson DG, Glass JI, Lartigue C, et al: Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 2010;329:52.
- Kornberg A: Basic research: The lifeline of medicine. *FASEB J* 1992;6:3143.
- Kornberg A: Centenary of the birth of modern biochemistry. *FASEB J* 1997;11:1209.
- Online Mendelian Inheritance in Man (OMIM): Center for Medical Genetics, Johns Hopkins University and National Center for Biotechnology Information, National Library of Medicine, 1997. <http://www.ncbi.nlm.nih.gov/omim/> (The numbers assigned to the entries in OMIM will be cited in selected chapters of this work. Consulting this extensive collection of diseases and other relevant entries—specific proteins, enzymes, etc—will greatly expand the reader's knowledge and understanding of various topics referred to and discussed in this text. The online version is updated almost daily.)
- Scriver CR, Beaudet AL, Valle D, et al (editors): *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed. McGraw-Hill, 2001 (This text is now available online and updated as *The Online Metabolic & Molecular Bases of Inherited Disease* at [www.ommbid.com](http://www.ommbid.com). Subscription is required, although access may be available via university and hospital libraries and other sources).
- Scherer S: *A Short Guide to the Human Genome*. CSHL Press, 2008.
- Weatherall DJ: Systems biology and red cells. *N Engl J Med* 2011;364;376.

## GLOSSARY

**Bioengineering:** The application of engineering to biology and medicine.

**Bioethics:** The area of ethics that is concerned with the application of moral and ethical principles to biology and medicine.

**Bioinformatics:** The discipline concerned with the collection, storage, and analysis of biologic data, mainly DNA and protein sequences (see Chapter 10).

**Biophysics:** The application of physics and its technics to biology and medicine.

**Biotechnology:** The field in which biochemical, engineering, and other approaches are combined to develop biological products of use in medicine and industry.

**Gene Therapy:** Applies to the use of genetically engineered genes to treat various diseases (see Chapter 39).

**Genomics:** The genome is the complete set of genes of an organism (eg, the human genome) and genomics is the in-depth study of the structures and functions of genomes (see Chapter 10 and other chapters).

**Glycomics:** The glycome is the total complement of simple and complex carbohydrates in an organism. Glycomics is the systematic study of the structures and functions of glycans (eg, the human glycome; see Chapter 47).

**Lipidomics:** The lipidome is the complete complement of lipids found in an organism. Lipidomics is the in-depth study of the structures and functions of all members of the lipidome and of their interactions, in both health and disease.

**Metabolomics:** The metabolome is the complete complement of metabolites (small molecules involved in metabolism) found in an organism. Metabolomics is the in-depth study of their structures, functions, and changes in various metabolic states.

**Molecular Diagnostics:** The use of molecular approaches (eg, DNA probes) to assist in the diagnosis of various biochemical, genetic, immunologic, microbiologic, and other medical conditions.

**Nanotechnology:** The development and application to medicine and to other areas of devices (such as nanoshells, see Glossary of Chapter 55) which are only a few nanometers in size. ( $10^{-9}$  m = 1 nm).

**Nutrigenomics:** The systematic study of the effects of nutrients on genetic expression and also of the effects of genetic variations on the handling of nutrients.

**Pharmacogenomics:** The use of genomic information and technologies to optimize the discovery and development of drug targets and drugs (see Chapter 54).

**Proteomics:** The proteome is the complete complement of proteins of an organism. Proteomics is the systematic study of the structures and functions of proteomes, including variations in health and disease (see Chapter 4).

**Stem Cell Biology:** A stem cell is an undifferentiated cell that has the potential to renew itself and to differentiate into any of the adult cells found in the organism. Stem cell biology is concerned with the biology of stem cells and their uses in various diseases.

**Synthetic Biology:** The field that combines biomolecular technics with engineering approaches to build new biological functions and systems.

**Systems Biology:** The field of science in which complex biologic systems are studied as integrated wholes (as opposed to the reductionist approach of, eg, classic biochemistry).

**Transcriptomics:** The transcriptome is the complete set of RNA transcripts produced by the genome at a fixed period in time. Transcriptomics is the comprehensive study of gene expression at the RNA level (see Chapter 36 and other chapters).

## OBJECTIVES

After studying this chapter, you should be able to:

- Describe the properties of water that account for its surface tension, viscosity, liquid state at ambient temperature, and solvent power.
- Use structural formulas to represent several organic compounds that can serve as hydrogen bond donors or acceptors.
- Explain the role played by entropy in the orientation, in an aqueous environment, of the polar and nonpolar regions of macromolecules.
- Indicate the quantitative contributions of salt bridges, hydrophobic interactions, and van der Waals forces to the stability of macromolecules.
- Explain the relationship of pH to acidity, alkalinity, and the quantitative determinants that characterize weak and strong acids.
- Calculate the shift in pH that accompanies the addition of a given quantity of acid or base to the pH of a buffered solution.
- Describe what buffers do, how they do it, and the conditions under which a buffer is most effective under physiologic or other conditions.
- Illustrate how the Henderson-Hasselbalch equation can be used to calculate the net charge on a polyelectrolyte at a given pH.

## BIOMEDICAL IMPORTANCE

Water is the predominant chemical component of living organisms. Its unique physical properties, which include the ability to solvate a wide range of organic and inorganic molecules, derive from water's dipolar structure and exceptional capacity for forming hydrogen bonds. The manner in which water interacts with a solvated biomolecule influences the structure both of the biomolecule and of water itself. An excellent nucleophile, water is a reactant or product in many metabolic reactions. Regulation of water balance depends upon hypothalamic mechanisms that control thirst, on antidiuretic hormone (ADH), on retention or excretion of water by the kidneys, and on evaporative loss. Nephrogenic diabetes insipidus, which involves the inability to concentrate urine or adjust to subtle changes in extracellular fluid osmolarity, results from the unresponsiveness of renal tubular osmoreceptors to ADH.

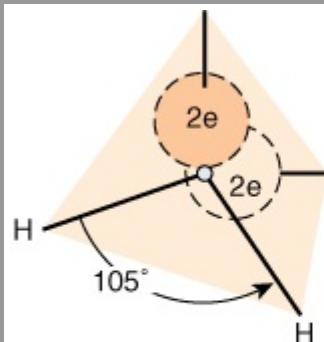
Water has a slight propensity to dissociate into hydroxide ions and protons. The concentration of protons, or **acidity**, of aqueous solutions is generally reported using the logarithmic pH scale. Bicarbonate and other buffers normally maintain the pH of extracellular fluid between 7.35 and 7.45. Suspected disturbances of acid-base balance are verified by measuring the pH of arterial blood and the CO<sub>2</sub> content of venous blood. Causes of acidosis (blood pH < 7.35) include diabetic ketosis and lactic acidosis. Alkalosis (pH > 7.45) may follow vomiting of acidic gastric contents.

## WATER IS AN IDEAL BIOLOGIC SOLVENT

### Water Molecules Form Dipoles

A water molecule is an irregular, slightly skewed tetrahedron with oxygen at its center (**Figure 2–1**). The two hydrogens and the unshared electrons of the remaining two  $sp^3$ -hybridized orbitals occupy the corners of the tetrahedron. The  $105^\circ$  angle between the hydrogen differs slightly from the ideal tetrahedral angle,  $109.5^\circ$ . Ammonia is also tetrahedral, with a  $107^\circ$  angle between its hydrogens. The strongly electronegative oxygen atoms in water attract electrons away from the hydrogen nuclei, leaving them with a partial positive charge, while its two unshared electron pairs constitute a region of local negative charge.

**Figure 2–1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

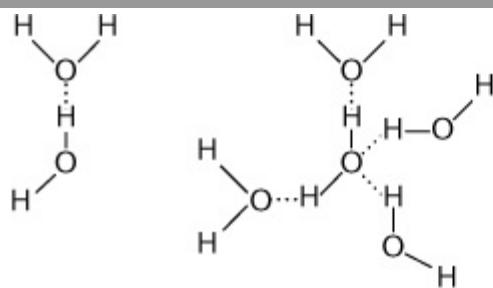
The water molecule has tetrahedral geometry.

A molecule with electrical charge distributed asymmetrically about its structure is referred to as a **dipole**. Water's strong dipole is responsible for its high **dielectric constant**. As described quantitatively by Coulomb's law, the strength of interaction  $F$  between oppositely charged particles is inversely proportionate to the dielectric constant  $\epsilon$  of the surrounding medium. The dielectric constant for a vacuum is unity; for hexane it is 1.9; for ethanol, 24.3; and for water, 78.5. Water therefore greatly decreases the force of attraction between charged and polar species relative to water-free environments with lower dielectric constants. Its strong dipole and high dielectric constant enable water to dissolve large quantities of charged compounds such as salts.

### Water Molecules Form Hydrogen Bonds

A partially unshielded hydrogen nucleus covalently bound to an electron-withdrawing oxygen or nitrogen atom can interact with an unshared electron pair on another oxygen or nitrogen atom to form a **hydrogen bond**. Since water molecules contain both of these features, hydrogen bonding favors the self-association of water molecules into ordered arrays (**Figure 2–2**). Hydrogen bonding profoundly influences the physical properties of water and accounts for its exceptionally high viscosity, surface tension, and boiling point. On average, each molecule in liquid water associates through hydrogen bonds with 3.5 others. These bonds are both relatively weak and transient, with a half-life of a few nanoseconds or less. Rupture of a hydrogen bond in liquid water requires only about 4.5 kcal/mol, less than 5% of the energy required to rupture a covalent O–H bond.

**Figure 2–2**



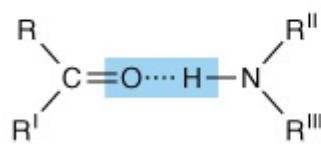
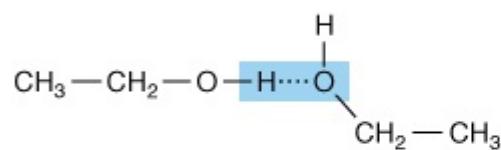
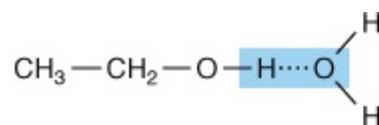
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Left:** Association of two dipolar water molecules by a hydrogen bond (dotted line). **Right:** Hydrogen-bonded cluster of four water molecules. Note that water can serve simultaneously both as a hydrogen donor and as a hydrogen acceptor.

Hydrogen bonding enables water to dissolve many organic biomolecules that contain functional groups which can participate in hydrogen bonding. The oxygen atoms of aldehydes, ketones, and amides, for example, provide lone pairs of electrons that can serve as hydrogen acceptors. Alcohols, carboxylic acids, and amines can serve both as hydrogen acceptors and as donors of unshielded hydrogen atoms for formation of hydrogen bonds (**Figure 2–3**).

**Figure 2–3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Additional polar groups participate in hydrogen bonding.** Shown are hydrogen bonds formed between alcohol and water, between two molecules of ethanol, and between the peptide carbonyl oxygen and the peptide nitrogen hydrogen of an adjacent amino acid.

## INTERACTION WITH WATER INFLUENCES THE STRUCTURE OF BIOMOLECULES

### Covalent and Noncovalent Bonds Stabilize Biologic Molecules

The covalent bond is the strongest force that holds molecules together (**Table 2–1**). Noncovalent forces, while of lesser magnitude, make significant contributions to the structure, stability, and functional competence of macromolecules in living cells. These forces, which can be either attractive or repulsive, involve interactions both within the biomolecule and between it and the water that forms the principal component of the surrounding environment.

**Table 2–1 Bond Energies for Atoms of Biologic Significance**

Bond Type	Energy (kcal/mol)	Bond Type	Energy (kcal/mol)
O—O	34	O=O	96
S—S	51	C—H	99
C—N	70	C=S	108
S—H	81	O—H	110
C—C	82	C=C	147
C—O	84	C=N	147
N—H	94	C=O	164

### Biomolecules Fold to Position Polar & Charged Groups on Their Surfaces

Most biomolecules are **amphipathic**; that is, they possess regions rich in charged or polar functional groups as well as regions with hydrophobic character. Proteins tend to fold with the R-groups of amino acids with hydrophobic side chains in the interior. Amino acids with charged or polar amino acid side chains (eg, arginine, glutamate, serine) generally are present on the surface in contact with water. A similar pattern prevails in a phospholipid bilayer, where the charged "head groups" of phosphatidyl serine or phosphatidyl ethanolamine contact water while their hydrophobic fatty acyl side chains cluster together, excluding water. This pattern maximizes the opportunities for the formation of energetically favorable charge-dipole, dipole-dipole, and hydrogen bonding interactions between polar groups on the biomolecule and water. It also minimizes energetically unfavorable contacts between water and hydrophobic groups.

### Hydrophobic Interactions

Hydrophobic interaction refers to the tendency of nonpolar compounds to self-associate in an aqueous environment. This self-association is driven neither by mutual attraction nor by what are sometimes incorrectly referred to as "hydrophobic bonds." Self-association minimizes the disruption of energetically favorable interactions between the surrounding water molecules.

While the hydrogens of nonpolar groups such as the methylene groups of hydrocarbons do not form hydrogen bonds, they do affect the structure of the water that surrounds them. Water molecules adjacent to a hydrophobic group are restricted in the number of orientations (degrees of freedom) that permit them to participate in the maximum number of energetically favorable hydrogen bonds. Maximal formation of multiple hydrogen bonds, which maximizes enthalpy, can be maintained only by increasing the order of the adjacent water molecules, with an accompanying decrease in entropy.

It follows from the second law of thermodynamics that the optimal free energy of a hydrocarbon–water mixture is a function of both maximal enthalpy (from hydrogen bonding) and minimum entropy (maximum degrees of freedom). Thus, nonpolar molecules tend to form droplets that minimize exposed surface area and reduce the number of water molecules whose motional freedom becomes restricted. Similarly, in the aqueous environment of the living cell the hydrophobic portions of biopolymers tend to be buried inside the structure of the molecule, or within a lipid bilayer, minimizing contact with water.

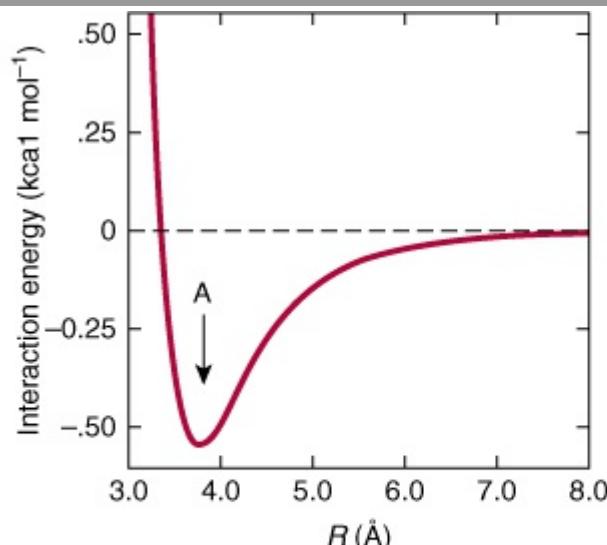
### Electrostatic Interactions

Interactions between charged groups help shape biomolecular structure. Electrostatic interactions between oppositely charged groups within or between biomolecules are termed **salt bridges**. Salt bridges are comparable in strength to hydrogen bonds but act over larger distances. They therefore often facilitate the binding of charged molecules and ions to proteins and nucleic acids.

### van der Waals Forces

van der Waals forces arise from attractions between transient dipoles generated by the rapid movement of electrons of all neutral atoms. Significantly weaker than hydrogen bonds but potentially extremely numerous, van der Waals forces decrease as the sixth power of the distance separating atoms (**Figure 2–4**). Thus, they act over very short distances, typically 2–4 Å.

**Figure 2–4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**The strength of van der Waals interactions varies with the distance, *R*, between interacting species.** The force of interaction between interacting species increases with decreasing distance until they are separated by the van der Waals contact distance (see arrow marked A). Repulsion due to interaction between the electrons of each atom or molecule then supervenes. While individual van der Waals interactions are extremely weak, the cumulative effect is nevertheless substantial for macromolecules such as DNA and proteins with many atoms in close contact.

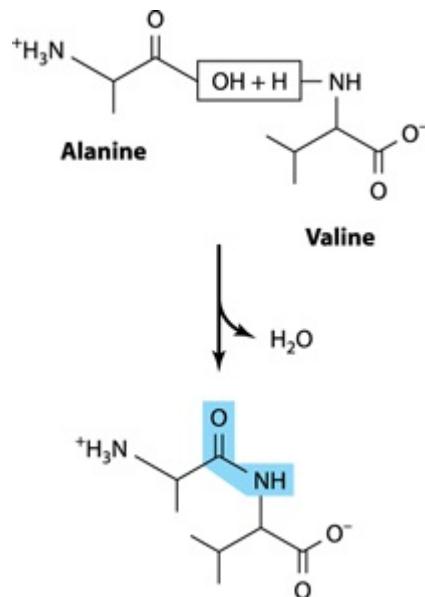
### Multiple Forces Stabilize Biomolecules

The DNA double helix illustrates the contribution of multiple forces to the structure of biomolecules. While each individual DNA strand is held together by covalent bonds, the two strands of the helix are held together exclusively by noncovalent interactions such as hydrogen bonds between nucleotide bases (Watson–Crick base pairing) and van der Waals interactions between the stacked purine and pyrimidine bases. The double helix presents the charged phosphate groups and polar hydroxyl groups from the ribose sugars of the DNA backbone to water while burying the relatively hydrophobic nucleotide bases inside. The extended backbone maximizes the distance between negatively charged phosphates, minimizing unfavorable electrostatic interactions.

## WATER IS AN EXCELLENT NUCLEOPHILE

Metabolic reactions often involve the attack by lone pairs of electrons residing on electron-rich molecules termed **nucleophiles** upon electron-poor atoms called **electrophiles**. Nucleophiles and electrophiles do not necessarily possess a formal negative or positive charge. Water, whose two lone pairs of  $sp^3$  electrons bear a partial negative charge (Figure 2–1), is an excellent nucleophile. Other nucleophiles of biologic importance include the oxygen atoms of phosphates, alcohols, and carboxylic acids; the sulfur of thiols; and the nitrogen of amines and the imidazole ring of histidine. Common electrophiles include the carbonyl carbons in amides, esters, aldehydes, and ketones and the phosphorus atoms of phosphoesters.

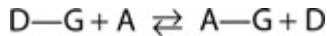
Nucleophilic attack by water typically results in the cleavage of the amide, glycoside, or ester bonds that hold biopolymers together. This process is termed **hydrolysis**. Conversely, when monomer units are joined together to form biopolymers such as proteins or glycogen, water is a product, for example, during the formation of a peptide bond between two amino acids:



While hydrolysis is a thermodynamically favored reaction, the amide and phosphoester bonds of polypeptides and oligonucleotides are stable in the aqueous environment of the cell. This seemingly paradoxical behavior reflects the fact that the thermodynamics governing the equilibrium of a reaction do not determine the rate at which it will proceed. In the cell, protein catalysts called **enzymes** accelerate the rate of hydrolytic reactions when needed. **Proteases** catalyze the hydrolysis of proteins into their component amino acids, while **nucleases** catalyze the hydrolysis of the phosphoester bonds in DNA and RNA. Careful control of the activities of these enzymes is required to ensure that they act only on appropriate target molecules at appropriate times.

## Many Metabolic Reactions Involve Group Transfer

Many of the enzymic reactions responsible for synthesis and breakdown of biomolecules involve the transfer of a chemical group G from a donor D to an acceptor A to form an acceptor group complex, A–G:



The hydrolysis and phosphorolysis of glycogen, for example, involve the transfer of glucosyl groups to water or to orthophosphate. The equilibrium constant for the hydrolysis of covalent bonds strongly favors the formation of split products. Conversely, in many cases the group transfer reactions responsible for the biosynthesis of macromolecules involve the thermodynamically unfavored formation of covalent bonds. Enzyme catalysts play a critical role in surmounting these barriers by virtue of their capacity to directly link two normally separate reactions together. By marrying an energetically unfavorable group transfer reaction with a thermodynamically favorable reaction, such as the hydrolysis of ATP, a new coupled reaction can be generated whose net *overall* change in free energy favors biopolymer synthesis.

Given the nucleophilic character of water and its high concentration in cells, why are biopolymers such as proteins and DNA relatively stable? And how can synthesis of biopolymers occur in an aqueous, seemingly prohydrolytic, environment? Central to both questions are the properties of enzymes. In the absence of enzymic catalysis, even reactions that are highly favored thermodynamically do not necessarily take place rapidly. Precise and differential control of enzyme activity and the sequestration of enzymes in specific organelles determine under what physiologic conditions a given biopolymer will be synthesized or degraded. Newly synthesized biopolymers are not immediately hydrolyzed because the active sites of biosynthetic enzymes sequester substrates in an environment from which water can be excluded.

## Water Molecules Exhibit a Slight But Important Tendency to Dissociate

The ability of water to ionize, while slight, is of central importance for life. Since water can act both as an acid and as a base, its ionization may be represented as an intermolecular proton transfer that forms a hydronium ion ( $H_3O^+$ ) and a hydroxide ion ( $OH^-$ ):



The transferred proton is actually associated with a cluster of water molecules. Protons exist in solution not only as  $H_3O^+$ , but also as multimers such as  $H_5O_2^+$  and  $H_7O_3^+$ . The proton is nevertheless routinely represented as  $H^+$ , even though it is in fact highly hydrated.

Since hydronium and hydroxide ions continuously recombine to form water molecules, an *individual* hydrogen or oxygen cannot be stated to be present as an ion or as part of a water molecule. At one instant it is an ion; an instant later it is part of a water molecule. Individual ions or molecules are therefore not considered. We refer instead to the *probability* that at any instant in time a given hydrogen will be present as an ion or as part of a water molecule. Since 1 g of water contains  $3.46 \times 10^{22}$  molecules, the ionization of water can be described statistically. To state that the probability that a hydrogen exists as an ion is 0.01 means that at any given moment in time, a hydrogen atom has 1 chance in 100 of being an ion and 99 chances out of 100 of being part of a water molecule. The actual probability of a hydrogen atom in pure water existing as a hydrogen ion is approximately  $1.8 \times 10^{-9}$ . The probability of its being part of a water molecule thus is almost unity. Stated another way, for every hydrogen ion or hydroxide ion in pure water, there are 1.8 billion or  $1.8 \times 10^9$  water molecules. Hydrogen ions and hydroxide ions nevertheless contribute significantly to the properties of water.

For dissociation of water,

$$K = \frac{[H^+][OH^-]}{[H_2O]}$$

where the brackets represent molar concentrations (strictly speaking, molar activities) and K is the **dissociation constant**. Since 1 mole (mol) of water weighs 18 g, 1 liter (L) (1000 g) of water contains  $1000 / 18 = 55.56$  mol. Pure water thus is 55.56 molar. Since the probability that a hydrogen in pure water will exist as a hydrogen ion is  $1.8 \times 10^{-9}$ , the molar concentration of  $H^+$  ions (or of  $OH^-$  ions) in pure water is the product of the probability,  $1.8 \times 10^{-9}$ , times the molar concentration of water, 55.56 mol/L. The result is  $1.0 \times 10^{-7}$  mol/L.

We can now calculate K for pure water:

$$K = \frac{[H^+][OH^-]}{[H_2O]} = \frac{[10^{-7}][10^{-7}]}{[55.56]} \\ = 0.018 \times 10^{-14} = 1.8 \times 10^{-16} \text{ mol/L}$$

The molar concentration of water, 55.56 mol/L, is too great to be significantly affected by dissociation. It is therefore considered to be essentially constant. This constant may therefore be incorporated into the dissociation constant K to provide a useful new constant  $K_w$  termed the **ion product** for water. The relationship between  $K_w$  and K is shown below:

$$K = \frac{[\text{H}^+][\text{OH}^-]}{[\text{H}_2\text{O}]} = 1.8 \times 10^{-16} \text{ mol/L}$$

$$\begin{aligned}K_w &= (K)[\text{H}_2\text{O}] = [\text{H}^+][\text{OH}^-] \\&= (1.8 \times 10^{-16} \text{ mol/L})(55.56 \text{ mol/L}) \\&= 1.00 \times 10^{-14} (\text{mol/L})^2\end{aligned}$$

Note that the dimensions of  $K$  are moles per liter and those of  $K_w$  are moles<sup>2</sup> per liter<sup>2</sup>. As its name suggests, the ion product  $K_w$  is numerically equal to the product of the molar concentrations of  $\text{H}^+$  and  $\text{OH}^-$ :

$$K_w = [\text{H}^+][\text{OH}^-]$$

At 25°C,  $K_w = (10^{-7})^2$ , or  $10^{-14}$  (mol/L)<sup>2</sup>. At temperatures below 25°C,  $K_w$  is somewhat less than  $10^{-14}$ , and at temperatures above 25°C it is somewhat greater than  $10^{-14}$ . Within the stated limitations of the effect of temperature,  $K_w$  equals  $10^{-14}$  (mol/L)<sup>2</sup> for all aqueous solutions, even solutions of acids or bases. We use  $K_w$  to calculate the pH of acidic and basic solutions.

## PH IS THE NEGATIVE LOG OF THE HYDROGEN ION CONCENTRATION

The term **pH** was introduced in 1909 by Sørensen, who defined pH as the negative log of the hydrogen ion concentration:

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

This definition, while not rigorous, suffices for many biochemical purposes. To calculate the pH of a solution:

1. Calculate the hydrogen ion concentration  $[\text{H}^+]$ .
2. Calculate the base 10 logarithm of  $[\text{H}^+]$ .
3. pH is the negative of the value found in step 2.

For example, for pure water at 25°C,

$$\text{pH} = -\log [\text{H}^+] = -\log 10^{-7} = -(-7) = 7.0$$

This value is also known as the *power* (English), *puissant* (French), or *potenz* (German) of the exponent, hence the use of the term "p."

Low pH values correspond to high concentrations of  $\text{H}^+$  and high pH values correspond to low concentrations of  $\text{H}^+$ .

Acids are **proton donors** and bases are **proton acceptors**. **Strong acids** (eg, HCl,  $\text{H}_2\text{SO}_4$ ) completely dissociate into anions and protons even in strongly acidic solutions (low pH). **Weak acids** dissociate only partially in acidic solutions. Similarly, **strong bases** (eg, KOH, NaOH)—but not **weak bases**, eg:  $\text{Ca}(\text{OH})_2$ , are completely dissociated even at high pH. Many biochemicals are weak acids. Exceptions include phosphorylated intermediates, whose phosphoryl group contains two dissociable protons, the first of which is strongly acidic.

The following examples illustrate how to calculate the pH of acidic and basic solutions.

**Example 1:** What is the pH of a solution whose hydrogen ion concentration is  $3.2 \times 10^{-4}$  mol/L?

$$\begin{aligned}\text{pH} &= -\log [\text{H}^+] \\ &= -\log (3.2 \times 10^{-4}) \\ &= -\log (3.2) - \log (10^{-4}) \\ &= -0.5 + 4.0 \\ &= 3.5\end{aligned}$$

**Example 2:** What is the pH of a solution whose hydroxide ion concentration is  $4.0 \times 10^{-4}$  mol/L? We first define a quantity **pOH** that is equal to  $-\log [\text{OH}^-]$  and that may be derived from the definition of  $K_w$ :

$$K_w = [\text{H}^+][\text{OH}^-] = 10^{-14}$$

Therefore,

$$\log [\text{H}^+] + \log [\text{OH}^-] = \log 10^{-14}$$

or

$$\text{pH} + \text{pOH} = 14$$

To solve the problem by this approach:

$$\begin{aligned}[\text{OH}^-] &= 4.0 \times 10^{-4} \\ \text{pOH} &= -\log [\text{OH}^-] \\ &= -\log (4.0 \times 10^{-4}) \\ &= -\log (4.0) - \log (10^{-4}) \\ &= -0.60 + 4.0 \\ &= 3.4\end{aligned}$$

Now

$$\text{pH} = 14 - \text{pOH} = 14 - 3.4$$

$$= 10.6$$

The examples above illustrate how the logarithmic pH scale facilitates recording and comparing hydrogen ion concentrations that differ by orders of magnitude from one another, ie, 0.00032 M (pH 3.5) and 0.00000000025 M (pH 10.6).

**Example 3:** What are the pH values of (a)  $2.0 \times 10^{-2}$  mol/L KOH and of (b)  $2.0 \times 10^{-6}$  mol/L KOH? The  $\text{OH}^-$  arises from two sources, KOH and water. Since pH is determined by the total  $[\text{H}^+]$  (and pOH by the total  $[\text{OH}^-]$ ), both sources must be considered. In the first case (a), the contribution of water to the total  $[\text{OH}^-]$  is negligible. The same cannot be said for the second case (b):

	Concentration (mol/L)	
	(a)	(b)
Molarity of KOH	$2.0 \times 10^{-2}$	$2.0 \times 10^{-6}$
$[\text{OH}^-]$ from KOH	$2.0 \times 10^{-2}$	$2.0 \times 10^{-6}$
$[\text{OH}^-]$ from water	$1.0 \times 10^{-7}$	$1.0 \times 10^{-7}$
Total $[\text{OH}^-]$	$2.00001 \times 10^{-2}$	$2.1 \times 10^{-6}$

Once a decision has been reached about the significance of the contribution by water, pH may be calculated as above.

The above examples assume that the strong base KOH is completely dissociated in solution and that the concentration of  $\text{OH}^-$  ions was thus equal to that due to the KOH plus that present initially in the water. This assumption is valid for dilute solutions of strong bases or acids, but not for weak bases or acids. Since weak electrolytes dissociate only slightly in solution, we must use the **dissociation constant** to calculate the concentration of  $[\text{H}^+]$  (or  $[\text{OH}^-]$ ) produced by a given molarity of a weak acid (or base) before calculating total  $[\text{H}^+]$  (or total  $[\text{OH}^-]$ ) and subsequently pH.

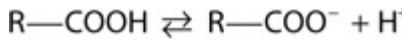
## Functional Groups that Are Weak Acids Have Great Physiologic Significance

Many biochemicals possess functional groups that are weak acids or bases. Carboxyl groups, amino groups, and phosphate esters, whose second dissociation falls within the physiologic range, are present in proteins and nucleic acids, most coenzymes, and most intermediary metabolites. Knowledge of the dissociation of weak acids and bases thus is basic to understanding the influence of intracellular pH on structure and biologic activity. Charge-based separations such as electrophoresis and ion exchange chromatography are also best understood in terms of the dissociation behavior of functional groups.

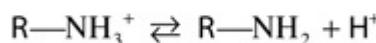
We term the protonated species (eg, HA or R-NH<sub>3</sub><sup>+</sup>) the **acid** and the unprotonated species (eg, A<sup>-</sup> or R-NH<sub>2</sub>) its **conjugate base**. Similarly, we may refer to a **base** (eg, A<sup>-</sup> or R-NH<sub>2</sub>) and its **conjugate acid** (eg, HA or R-NH<sub>3</sub><sup>+</sup>). Representative weak acids (left), their conjugate bases (center), and pK<sub>a</sub> values (right) include the following:

R—CH <sub>2</sub> —COOH	R—CH <sub>2</sub> —COO <sup>-</sup>	pK <sub>a</sub> = 4 – 5
R—CH <sub>2</sub> —NH <sub>3</sub> <sup>+</sup>	R—CH <sub>2</sub> —NH <sub>2</sub>	pK <sub>a</sub> = 9 – 10
H <sub>2</sub> CO <sub>3</sub>	HCO <sub>3</sub> <sup>-</sup>	pK <sub>a</sub> = 6.4
H <sub>2</sub> PO <sub>4</sub> <sup>-</sup>	HPO <sub>4</sub> <sup>-2</sup>	pK <sub>a</sub> = 7.2

We express the relative strengths of weak acids and bases in terms of their dissociation constants. Shown below are the expressions for the dissociation constant (K<sub>a</sub>) for two representative weak acids, R-COOH and R-NH<sub>3</sub><sup>+</sup>.



$$K_a = \frac{[\text{R}-\text{COO}^-][\text{H}^+]}{[\text{R}-\text{COOH}]}$$



$$K_a = \frac{[\text{R}-\text{NH}_2][\text{H}^+]}{[\text{R}-\text{NH}_3^+]}$$

Since the numeric values of K<sub>a</sub> for weak acids are negative exponential numbers, we express K<sub>a</sub> as pK<sub>a</sub>, where

$$\text{pK}_a = -\log K_a$$

Note that pK<sub>a</sub> is related to K<sub>a</sub> as pH is to [H<sup>+</sup>]. The stronger the acid, the lower is its pK<sub>a</sub> value.

pK<sub>a</sub> is used to express the relative strengths of both acids and bases. For any weak acid, its conjugate is a strong base. Similarly, the conjugate of a strong base is a weak acid. **The relative strengths of bases are expressed in terms of the pK<sub>a</sub> of their conjugate acids.** For polyprotic compounds containing more than one dissociable proton, a numerical subscript is assigned to each dissociation, numbered starting from unity in decreasing order of relative acidity. For a dissociation of the type



the pK<sub>a</sub> is the pH at which the concentration of the acid R-NH<sub>3</sub><sup>+</sup> equals that of the base R-NH<sub>2</sub>.

From the above equations that relate K<sub>a</sub> to [H<sup>+</sup>] and to the concentrations of undissociated acid and its conjugate base, when

$$[\text{R}-\text{COO}^-] = [\text{R}-\text{COOH}]$$

or when

$$[\text{R}-\text{NH}_2] = [\text{R}-\text{NH}_3^+]$$

then

$$K_a = [\text{H}^+]$$

Thus, when the associated (protonated) and dissociated (conjugate base) species are present at equal concentrations, the prevailing hydrogen ion concentration [H<sup>+</sup>] is numerically equal to the dissociation constant, K<sub>a</sub>. If the logarithms of both sides of the above equation are taken and both sides are multiplied by -1, the expressions would be as follows:

$$K_a = [\text{H}^+]$$

$$-\log K_a = -\log [\text{H}^+]$$

Since -log K<sub>a</sub> is defined as pK<sub>a</sub>, and -log [H<sup>+</sup>] defines pH, the equation may be rewritten as

$$\text{pK}_a = \text{pH}$$

ie, **the pK<sub>a</sub> of an acid group is the pH at which the protonated and unprotonated species are present at equal concentrations.** The pK<sub>a</sub> for an acid may be determined by adding 0.5 equivalent of alkali per equivalent of acid. The resulting pH will equal the pK<sub>a</sub> of the acid.

## The Henderson–Hasselbalch Equation Describes the Behavior of Weak Acids & Buffers

The Henderson–Hasselbalch equation is derived below.

A weak acid, HA, ionizes as follows:



The equilibrium constant for this dissociation is

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]}$$

Cross-multiplication gives

$$[\text{H}^+][\text{A}^-] = K_a[\text{HA}]$$

Divide both sides by [A<sup>-</sup>]:

$$[\text{H}^+] = K_a \frac{[\text{HA}]}{[\text{A}^-]}$$

Take the log of both sides:

$$\log[H^+] = \log \left( K_a \frac{[HA]}{[A^-]} \right)$$

$$= \log K_a + \log \frac{[HA]}{[A^-]}$$

Multiply through by -1:

$$-\log[H^+] = -\log K_a - \log \frac{[HA]}{[A^-]}$$

Substitute pH and  $pK_a$  for  $-\log[H^+]$  and  $-\log K_a$ , respectively; then

$$pH = pK_a - \log \frac{[HA]}{[A^-]}$$

Inversion of the last term removes the minus sign and gives the **Henderson-Hasselbalch equation**

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$

The Henderson-Hasselbalch equation has great predictive value in protonic equilibria. For example,

- When an acid is exactly half-neutralized,  $[A^-] = [HA]$ . Under these conditions,

$$pH = pK_a + \log \frac{[A^-]}{[HA]} = pK_a + \log \left( \frac{1}{1} \right) = pK_a + 0$$

Therefore, at half-neutralization,  $pH = pK_a$ .

- When the ratio  $[A^-]/[HA] = 100:1$ ,

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$

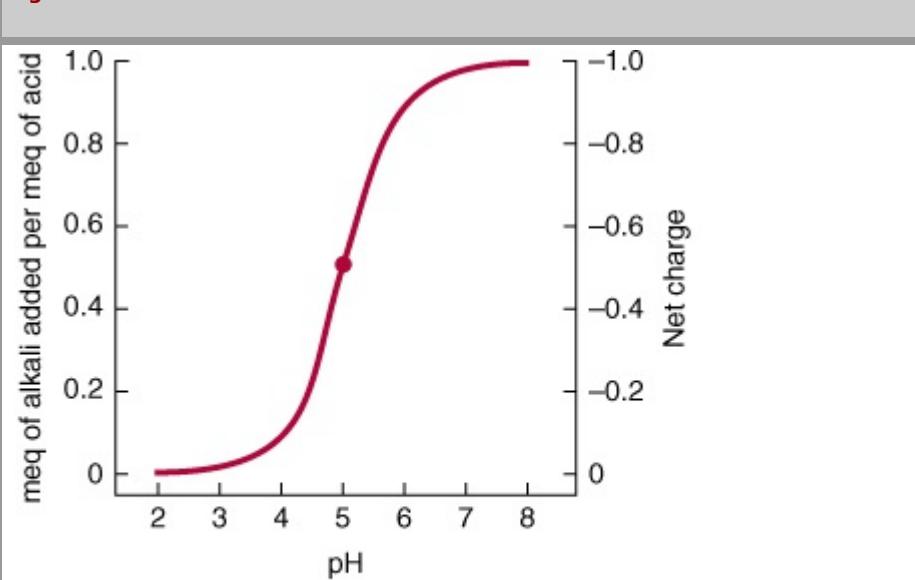
$$pH = pK_a + \log (100/1) = pK_a + 2$$

- When the ratio  $[A^-]/[HA] = 1:10$ ,

$$pH = pK_a + \log (1/10) = pK_a + (-1)$$

If the equation is evaluated at ratios of  $[A^-]/[HA]$  ranging from  $10^3$  to  $10^{-3}$  and the calculated pH values are plotted, the resulting graph describes the titration curve for a weak acid (**Figure 2-5**).

**Figure 2-5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Titration curve for an acid of the type HA. The heavy dot in the center of the curve indicates the  $pK_a$  5.0.

## Solutions of Weak Acids & Their Salts Buffer Changes in pH

Solutions of weak acids or bases and their conjugates exhibit **buffering**, the ability to resist a change in pH following addition of strong acid or base. Since many metabolic reactions are accompanied by the release or uptake of protons, most intracellular reactions are buffered. Oxidative metabolism produces  $\text{CO}_2$ , the anhydride of carbonic acid, which if not buffered would produce severe acidosis. Maintenance of a constant pH involves buffering by phosphate, bicarbonate, and proteins, which accept or release protons to resist a change in pH. For experiments using tissue extracts or enzymes, constant pH is maintained by the addition of buffers such as MES ([2-N-morpholino]ethanesulfonic acid,  $pK_a$  6.1), inorganic orthophosphate ( $pK_{a2}$  7.2), HEPES (*N*-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid,  $pK_a$  6.8), or Tris (tris[hydroxymethyl]aminomethane,  $pK_a$  8.3). The value of  $pK_a$  relative to the desired pH is the major determinant of which buffer is selected.

Buffering can be observed by using a pH meter while titrating a weak acid or base (**Figure 2–5**). We can also calculate the pH shift that accompanies addition of acid or base to a buffered solution. In the example, the buffered solution (a weak acid,  $pK_a = 5.0$ , and its conjugate base) is initially at one of four pH values. We will calculate the pH shift that results when 0.1 meq of KOH is added to 1 meq of each solution:

Initial pH	5.00	5.37	5.60	5.86
$[A^-]_{\text{initial}}$	0.50	0.70	0.80	0.88
$[\text{HA}]_{\text{initial}}$	0.50	0.30	0.20	0.12
$([A^-]/[\text{HA}])_{\text{initial}}$	1.00	2.33	4.00	7.33
<b>Addition of 0.1 meq of KOH produces</b>				
$[A^-]_{\text{final}}$	0.60	0.80	0.90	0.98
$[\text{HA}]_{\text{final}}$	0.40	0.20	0.10	0.02
$([A^-]/[\text{HA}])_{\text{final}}$	1.50	4.00	9.00	49.0
$\log ([A^-]/[\text{HA}])_{\text{final}}$	0.18	0.60	0.95	1.69
Final pH	5.18	5.60	5.95	6.69
<b>pH</b>	<b>0.18</b>	<b>0.60</b>	<b>0.95</b>	<b>1.69</b>

Notice that the change in pH per milliequivalent of  $\text{OH}^-$  added depends on the initial pH. The solution resists changes in pH most effectively at pH values close to the  $pK_a$ . **A solution of a weak acid and its conjugate base buffers most effectively in the pH range  $pK_a \pm 1.0$  pH unit.**

Figure 2–5 also illustrates the net charge on one molecule of the acid as a function of pH. A fractional charge of  $-0.5$  does not mean that an individual molecule bears a fractional charge but that the *probability* is 0.5 that a given molecule has a unit negative charge at any given moment in time. Consideration of the net charge on macromolecules as a function of pH provides the basis for separatory techniques such as ion exchange chromatography and electrophoresis.

## Acid Strength Depends on Molecular Structure

Many acids of biologic interest possess more than one dissociating group. The presence of adjacent negative charge hinders the release of a proton from a nearby group, raising its  $pK_a$ . This is apparent from the  $pK_a$  values for the three dissociating groups of phosphoric acid and citric acid (**Table 2–2**). The effect of adjacent charge decreases with distance. The second  $pK_a$  for succinic acid, which has two methylene groups between its carboxyl groups, is 5.6, whereas the second  $pK_a$  for glutaric acid, which has one additional methylene group, is 5.4.

**Table 2–2 Relative Strengths of Selected Acids of Biologic Significance<sup>1</sup>**

<b>Monoprotic Acids</b>		
Formic	$pK$	3.75
Lactic	$pK$	3.86
Acetic	$pK$	4.76
Ammonium ion	$pK$	9.25
<b>Diprotic Acids</b>		
Carbonic	$pK_1$	6.37
	$pK_2$	10.25
Succinic	$pK_1$	4.21
	$pK_2$	5.64
Glutaric	$pK_1$	4.34
	$pK_2$	5.41
<b>Triprotic Acids</b>		
Phosphoric	$pK_1$	2.15
	$pK_2$	6.82
	$pK_3$	12.38
Citric	$pK_1$	3.08
	$pK_2$	4.74
	$pK_3$	5.40

<sup>1</sup>**Note:** Tabulated values are the  $pK_a$  values ( $-\log$  of the dissociation constant) of selected monoprotic, diprotic, and triprotic acids.

## $pK_a$ Values Depend on the Properties of the Medium

The  $pK_a$  of a functional group is also profoundly influenced by the surrounding medium. The medium may either raise or lower the  $pK_a$  depending on whether the undissociated acid or its conjugate base is the charged species. The effect of dielectric constant on  $pK_a$  may be observed by adding ethanol to water. The  $pK_a$  of a carboxylic acid *increases*, whereas that of an amine *decreases* because ethanol decreases the ability of water to solvate a charged species. The  $pK_a$  values of dissociating groups in the interiors of proteins thus are profoundly affected by their local environment, including the presence or absence of water.

## SUMMARY

- Water forms hydrogen-bonded clusters with itself and with other proton donors or acceptors. Hydrogen bonds account for the surface tension, viscosity, liquid state at room temperature, and solvent power of water.
- Compounds that contain O or N can serve as hydrogen bond donors and/or acceptors.
- Macromolecules exchange internal surface hydrogen bonds for hydrogen bonds to water. Entropic forces dictate that macromolecules expose polar regions to an aqueous interface and bury nonpolar regions.
- Salt bridges, hydrophobic interactions, and van der Waals forces participate in maintaining molecular structure.
- pH is the negative log of  $[H^+]$ . A low pH characterizes an acidic solution, and a high pH denotes a basic solution.
- The strength of weak acids is expressed by  $pK_a$ , the negative log of the acid dissociation constant. Strong acids have low  $pK_a$  values and weak acids have high  $pK_a$  values.
- Buffers resist a change in pH when protons are produced or consumed. Maximum buffering capacity occurs  $\pm 1$  pH unit on either side of  $pK_a$ . Physiologic buffers include bicarbonate, orthophosphate, and proteins.

## REFERENCES

- Reese KM: Whence came the symbol pH. *Chem & Eng News* 2004;82:64.
- Segel IM: *Biochemical Calculations*. Wiley, 1968.
- Skinner JL: Following the motions of water molecules in aqueous solutions. *Science* 2010;328:985.
- Stillinger FH: Water revisited. *Science* 1980;209:451.
- Suresh SJ, Naik VM: Hydrogen bond thermodynamic properties of water from dielectric constant data. *J Chem Phys* 2000;113:9727.
- Wiggins PM: Role of water in some biological processes. *Microbiol Rev* 1990;54:432.

## OBJECTIVES

After studying this chapter, you should be able to:

- Name, and draw the structures of, the 20 amino acids present in proteins.
- Write the three- and one-letter designations for each of the common amino acids.
- List the ionizable groups of the common amino acids and their  $pK_a$  values.
- Calculate the pH of an unbuffered aqueous solution of a polyfunctional amino acid and the change in pH that occurs following the addition of a given quantity of strong acid or alkali.
- Define pI and indicate its relationship to the net charge on a polyfunctional electrolyte.
- Explain how pH,  $pK_a$  and pI can be used to predict the mobility of a polyelectrolyte, such as an amino acid, in a direct-current electrical field.
- Describe the contribution of each type of R group of the common amino acids to their chemical properties.
- Describe the directionality, nomenclature, and primary structure of peptides.
- Identify the bond in a peptide that exhibits partial double-bond character and its conformational consequences in a peptide.
- Identify those bonds in the peptide backbone that are capable of free rotation and the Greek letters used to designate them.

## BIOMEDICAL IMPORTANCE

In addition to providing the monomer units from which the long polypeptide chains of proteins are synthesized, the L- $\alpha$ -amino acids and their derivatives participate in cellular functions as diverse as nerve transmission and the biosynthesis of porphyrins, purines, pyrimidines, and urea. Short polymers of amino acids called *peptides* perform prominent roles in the neuroendocrine system as hormones, hormone-releasing factors, neuromodulators, or neurotransmitters. Humans and other higher animals lack the capability to synthesize 10 of the 20 common L- $\alpha$ -amino acids in amounts adequate to support infant growth or to maintain health in adults. Consequently, the human diet must contain adequate quantities of these nutritionally essential amino acids. While human proteins contain only L- $\alpha$ -amino acids, microorganisms make extensive use of D- $\alpha$ -amino acids. *Bacillus subtilis*, for example, secretes a mixture of D-methionine, D-tyrosine, D-leucine, and D-tryptophan to trigger biofilm disassembly, and *Vibrio cholerae* incorporates D-leucine and D-methionine into the peptide component of their peptidoglycan layer. Many bacteria elaborate peptides that contain both D- and L- $\alpha$ -amino acids, several of which possess therapeutic value, including the antibiotics bacitracin and gramicidin A and the antitumor agent bleomycin. Certain other microbial peptides are toxic. The cyanobacterial peptides microcystin and nodularin are lethal in large doses, while small quantities promote the formation of hepatic tumors.

## PROPERTIES OF AMINO ACIDS

### The Genetic Code Specifies 20 L-α-Amino Acids

Of the over 300 naturally occurring amino acids, 20 constitute the predominant monomer units of proteins. While a three-letter genetic code could potentially accommodate more than 20 amino acids, several amino acids are specified by multiple codons (see **Table 37-1**). Redundant usage limits the available codons to the 20 L-α-amino acids listed in **Table 3-1**. Both one- and three-letter abbreviations for each amino acid can be used to represent the amino acids in peptides and proteins (Table 3-1). Some proteins contain additional amino acids that are by modification of an amino acid already present in a peptide. Examples include conversion of peptidyl proline and lysine to 4-hydroxyproline and 5-hydroxylysine; the conversion of peptidyl glutamate to -carboxyglutamate; and the methylation, formylation, acetylation, prenylation, and phosphorylation of certain aminoacyl residues. These modifications extend the biologic diversity of proteins by altering their solubility, stability, and interaction with other proteins.

**Table 3-1 L-α-Amino Acids Present in Proteins**

Name	Symbol	Structural Formula	pK <sub>1</sub>	pK <sub>2</sub>	pK <sub>3</sub>
<b>With Aliphatic Side Chains</b>			-COOH	-NH <sub>3</sub> <sup>+</sup>	R Group
Glycine	Gly [G]	H—CH—COO <sup>−</sup>   NH <sub>3</sub> <sup>+</sup>	2.4	9.8	
Alanine	Ala [A]	CH <sub>3</sub> —CH—COO <sup>−</sup>   NH <sub>3</sub> <sup>+</sup>	2.4	9.9	
Valine	Val [V]	H <sub>3</sub> C   CH—CH—COO <sup>−</sup>   H <sub>3</sub> C      NH <sub>3</sub> <sup>+</sup>	2.2	9.7	
Leucine	Leu [L]	H <sub>3</sub> C   CH—CH <sub>2</sub> —CH—COO <sup>−</sup>   H <sub>3</sub> C      NH <sub>3</sub> <sup>+</sup>	2.3	9.7	
Isoleucine	Ile [I]	CH <sub>3</sub>   CH <sub>2</sub>   CH—CH—COO <sup>−</sup>   CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	2.3	9.8	
<b>With Side Chains Containing Hydroxylic (OH) Groups</b>					
Serine	Ser [S]	CH <sub>2</sub> —CH—COO <sup>−</sup>        OH    NH <sub>3</sub> <sup>+</sup>	2.2	9.2	about 13
Threonine	Thr [T]	CH <sub>3</sub> —CH—CH—COO <sup>−</sup>        OH    NH <sub>3</sub> <sup>+</sup>	2.1	9.1	about 13
Tyrosine	Tyr [Y]	See below.			
Cysteine	Cys [C]	CH <sub>2</sub> —CH—COO <sup>−</sup>        SH    NH <sub>3</sub> <sup>+</sup>	1.9	10.8	8.3
Methionine	Met [M]	CH <sub>2</sub> —CH <sub>2</sub> —CH—COO <sup>−</sup>   S—CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	2.1	9.3	
<b>With Side Chains Containing Acidic Groups or Their Amides</b>					
Aspartic acid	Asp [D]	—OOC—CH <sub>2</sub> —CH—COO <sup>−</sup>   NH <sub>3</sub> <sup>+</sup>	2.1	9.9	3.9
Asparagine	Asn [N]	H <sub>2</sub> N—C—CH <sub>2</sub> —CH—COO <sup>−</sup>         O    NH <sub>3</sub> <sup>+</sup>	2.1	8.8	
Glutamic acid	Glu [E]	—OOC—CH <sub>2</sub> —CH <sub>2</sub> —CH—COO <sup>−</sup>   NH <sub>3</sub> <sup>+</sup>	2.1	9.5	4.1
Glutamine	Gln [Q]	H <sub>2</sub> N—C—CH <sub>2</sub> —CH <sub>2</sub> —CH—COO <sup>−</sup>         O    NH <sub>3</sub> <sup>+</sup>	2.2	9.1	
<b>With Side Chains Containing Basic Groups</b>					
Arginine	Arg [R]	H—N—CH <sub>2</sub> —CH <sub>2</sub> —CH <sub>2</sub> —CH—COO <sup>−</sup>                        C=NH <sub>2</sub> <sup>+</sup> NH <sub>3</sub> <sup>+</sup>   NH <sub>2</sub>	1.8	9.0	12.5

Lysine	Lys [K]	The structure shows a primary amine group (NH3+) at the C-terminal end.	2.2	9.2	10.8
Histidine	His [H]	The structure shows a secondary amine group (NH3+) at the C-terminal end, attached to an imidazole ring.	1.8	9.3	6.0
<b>Containing Aromatic Rings</b>					
Histidine	His [H]	See above.			
Phenylalanine	Phe [F]	The structure shows a secondary amine group (NH3+) at the C-terminal end, attached to a phenyl ring.	2.2	9.2	
Tyrosine	Tyr [Y]	The structure shows a secondary amine group (NH3+) at the C-terminal end, attached to a hydroxylated phenyl ring.	2.2	9.1	10.1
Tryptophan	Trp [W]	The structure shows a secondary amine group (NH3+) at the C-terminal end, attached to an indole ring.	2.4	9.4	
<b>Imino Acid</b>					
Proline	Pro [P]	The structure shows a secondary amine group (NH3+) at the C-terminal end, part of a five-membered imidazolidine ring.	2.0	10.6	

## Selenocysteine, the 21st L-α-Amino Acid

Selenocysteine is an L-α-amino acid found in proteins from every domain of life. Humans contain approximately two dozen selenoproteins that include certain peroxidases and reductases selenoprotein P which circulates in the plasma, and the iodothyronine deiodinases responsible for converting the prohormone thyroxine (T4) to the thyroid hormone 3,3'-5-triiodothyronine (T3) (Chapter 41). As the name implies, a selenium atom replaces the sulfur of its structural analog, cysteine. The  $pK_3$  of selenocysteine, 5.2, is three units lower than that of cysteine. Unlike other unusual amino acids, selenocysteine is not the product of a posttranslational modification. Rather, it is inserted directly into a growing polypeptide during translation. Selenocysteine thus is commonly referred to as the "21st amino acid." However, unlike other 20 genetically encoded amino acids, selenocysteine is specified by a much larger and more complex genetic element than the basic three-letter codon (see Chapter 27).

## Only L-α-Amino Acids Occur in Proteins

With the sole exception of glycine, the α-carbon of every amino acid is chiral. Although some protein amino acids are dextrorotatory and some levorotatory, all share the absolute configuration of L-glyceraldehyde and thus are defined as L-α-amino acids. Several free L-α-amino acids fulfill important roles in metabolic processes. Examples include ornithine, citrulline, and argininosuccinate that participate in urea synthesis; tyrosine in formation of thyroid hormones; and glutamate in neurotransmitter biosynthesis. D-amino acids that occur naturally include free D-serine and D-aspartate in brain tissue, D-alanine and D-glutamate in the cell walls of gram-positive bacteria, and D-amino acids in certain peptides and antibiotics produced by bacteria, fungi, reptiles, and other nonmammalian species.

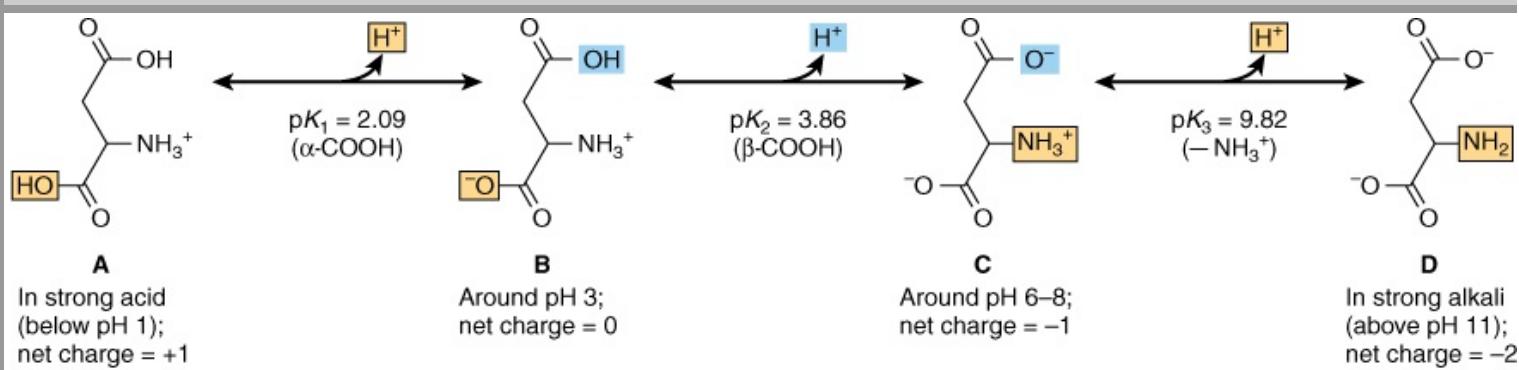
## Amino Acids May Have Positive, Negative, or Zero Net Charge

Charged and uncharged forms of the ionizable  $-COOH$  and  $-NH_3^+$  weak acid groups that exist in solution in protonic equilibrium:



While both  $R-COOH$  and  $R-NH_3^+$  are weak acids,  $R-COOH$  is a far stronger acid than  $R-NH_3^+$ . Thus, at physiologic pH (pH 7.4), carboxyl groups exist almost entirely as  $R-COO^-$  and amino groups predominantly as  $R-NH_3^+$ . **Figure 3-1** illustrates the effect of pH on the charged state of aspartic acid.

**Figure 3-1**

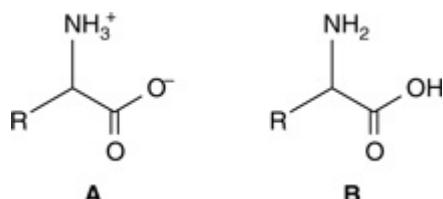


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Protonic equilibria of aspartic acid.**

Molecules that contain an equal number of ionizable groups of opposite charge and that therefore bear no net charge are termed **zwitterions**. Amino acids in blood and most tissues thus should be represented as in **A**, below.

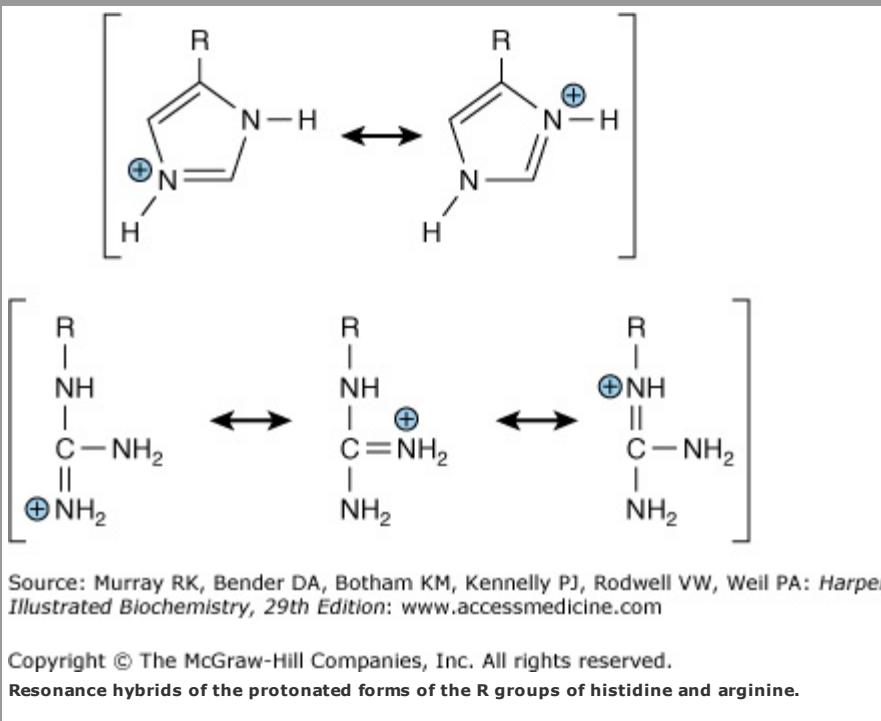


Structure **B** cannot exist in an aqueous solution because at any pH low enough to protonate the carboxyl group, the amino group would also be protonated. Similarly, at any pH sufficient high for an uncharged amino group to predominate, a carboxyl group will be present as R-COO<sup>-</sup>. The uncharged representation **B** is, however, often used for reactions that do not involve protonic equilibria.

### PK<sub>a</sub> Values Express the Strengths of Weak Acids

The acid strengths of weak acids are expressed as their pK<sub>a</sub>. For molecules with multiple dissociable protons, the pK<sub>a</sub> for each acidic group is designated by replacing the subscript "a" with a number (Table 3-1). The imidazole group of histidine and the guanidino group of arginine exist as resonance hybrids with positive charge distributed between both nitrogens (histidine) or all three nitrogens (arginine) (Figure 3-2). The net charge on an amino acid—the algebraic sum of all the positively and negatively charged groups present—depends upon the pK<sub>a</sub> values of its functional groups and on the pH of the surrounding medium. Altering the charge on amino acids and their derivatives by varying the pH facilitates the physical separation of amino acids, peptides, and proteins (see Chapter 4).

**Figure 3-2**



### At Its Isoelectric pH (pI), an Amino Acid Bears No Net Charge

Zwitterions are one example of an **isoelectric** species—the form of a molecule that has an equal number of positive and negative charges and thus is electrically neutral. The isoelectric pH, also called the pI, is the pH midway between pK<sub>a</sub> values for the ionizations on either side of the isoelectric species. For an amino acid such as alanine that has only two dissociating groups, there is no ambiguity. The first pK<sub>a</sub> (R-COOH) is 2.35 and the second pK<sub>a</sub> (R-NH<sub>3</sub><sup>+</sup>) is 9.69. The isoelectric pH (pI) of alanine thus is

$$pI = \frac{pK_1 + pK_2}{2} = \frac{2.35 + 9.69}{2} = 6.02$$

For polyprotic acids, pI is also the pH midway between the pK<sub>a</sub> values on either side of the isoionic species. For example, the pI for aspartic acid is

$$pI = \frac{pK_1 + pK_2}{2} = \frac{2.09 + 3.96}{2} = 3.02$$

For lysine, pI is calculated from

$$pI = \frac{pK_2 + pK_3}{2}$$

Similar considerations apply to all polyprotic acids (eg, proteins), regardless of the number of dissociating groups present. In the clinical laboratory, knowledge of the pI guides selection of conditions for electrophoretic separations. For example, electrophoresis at pH 7.0 will separate two molecules with pI values of 6.0 and 8.0, because at pH 7.0 the molecule with a pI of 6.0 will have a net positive charge, and that with a pI of 8.0 a net negative charge. Similar considerations apply to understanding chromatographic separations on ionic supports such as diethylaminoethyl (DEAE) cellulose (see Chapter 4).

### PK<sub>a</sub> Values Vary with the Environment

The environment of a dissociable group affects its pK<sub>a</sub>. The pK<sub>a</sub> values of the R groups of free amino acids in an aqueous solution (Table 3-1) thus provide only an approximate guide to the pK<sub>a</sub> values of the same amino acids when present in proteins. A polar environment favors the charged form (R-COO<sup>-</sup> or R-NH<sub>3</sub><sup>+</sup>), and a nonpolar environment favors the uncharged form (R-COOH or R-NH<sub>2</sub>). A nonpolar environment thus *raises* the pK<sub>a</sub> of a carboxyl group (making it a weaker acid) but *lowers* that of an amino group (making it a stronger acid). The presence of adjacent charged groups can reinforce or counteract solvent effects. The pK<sub>a</sub> of a functional group thus will depend upon its location within a given protein. Variations in pK<sub>a</sub> can encompass whole pH units (Table 3-2). pK<sub>a</sub> values that diverge from those listed by as much as 3 pH units are common at the active sites of enzymes. An extreme example, a buried aspartic acid of thioredoxin, has a pK<sub>a</sub> above 9—a shift of more than 6 pH units.

**Table 3-2 Typical Range of PK<sub>a</sub> Typical Range of PK<sub>a</sub> Values for Ionizable Groups in Proteins**

Dissociating Group	pK <sub>a</sub> Range
-Carboxyl	3.5–4.0
Non- <sup>-</sup> COOH of Asp or Glu	4.0–4.8
Imidazole of His	6.5–7.4
SH of Cys	8.5–9.0

OH of Tyr	9.5–10.5
-Amino	8.0–9.0
-Amino of Lys	9.8–10.4
Guanidinium of Arg	~12.0

## The Solubility of Amino Acids Reflects Their Ionic Character

The charges conferred by the dissociable functional groups of amino acids ensure that they are readily solvated by—and thus soluble in—polar solvents such as water and ethanol but insoluble in nonpolar solvents such as benzene, hexane, or ether.

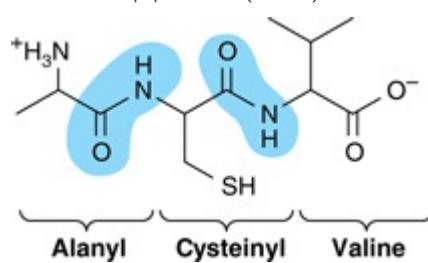
Amino acids do not absorb visible light and thus are colorless. However, tyrosine, phenylalanine, and especially tryptophan absorb high-wavelength (250–290 nm) ultraviolet light. Because it absorbs ultraviolet light about ten times more efficiently than phenylalanine or tyrosine, tryptophan makes the major contribution to the ability of most proteins to absorb light in the region of 280 nm.

## THE $\alpha$ -R GROUPS DETERMINE THE PROPERTIES OF AMINO ACIDS

Since glycine, the smallest amino acid, can be accommodated in places inaccessible to other amino acids, it often occurs where peptides bend sharply. The hydrophobic R groups of alanine, valine, leucine, and isoleucine and the aromatic R groups of phenylalanine, tyrosine, and tryptophan typically occur primarily in the interior of cytosolic proteins. The charged R groups of basic and acidic amino acids stabilize specific protein conformations via ionic interactions or salt bridges. These interactions also function in "charge relay" systems during enzymatic catalysis and electron transport in respiring mitochondria. Histidine plays unique roles in enzymatic catalysis. The  $pK_a$  of its imidazole proton permits histidine to function at neutral pH as either a base or an acid catalyst without the need for any environmentally induced shift. The primary alcohol group of serine and the primary thioalcohol ( $-SH$ ) group of cysteine are excellent nucleophiles, and can function as such during enzymatic catalysis. However, the secondary alcohol group of threonine, while a good nucleophile, is not known to fulfill an analogous role in catalysis. The  $-OH$  groups of serine, tyrosine, and threonine also participate in regulation of the activity of enzymes whose catalytic activity depends on the phosphorylation state of these residues.

## FUNCTIONAL GROUPS DICTATE THE CHEMICAL REACTIONS OF AMINO ACIDS

Each functional group of an amino acid exhibits all of its characteristic chemical reactions. For carboxylic acid groups, these reactions include the formation of esters, amides, and acid anhydrides; for amino groups, acylation, amidation, and esterification; and for  $-\text{OH}$  and  $-\text{SH}$  groups, oxidation, and esterification. The most important reaction of amino acids is the formation of a peptide bond (shaded).

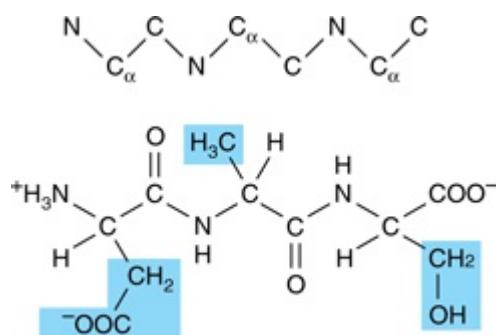


### Amino Acid Sequence Determines Primary Structure

The number and order of all the amino acid residues in a polypeptide constitute its primary structure. Amino acids present in peptides are called aminoacyl residues and are named by replacing the *-ate* or *-ine* suffixes of free amino acids with *-yl* (eg, alanyl, aspartyl, tyrosyl). Peptides are then named as derivatives of the carboxy terminal aminoacyl residue. For example, Lys-Leu-Tyr-Gln is called lysyl-leucyl-tyrosyl-glutamine. The *-ine* ending on glutamine indicates that its  $-\text{COO}^-$  group is *not* involved in peptide bond formation.

### Peptide Structures Are Easy to Draw

Prefixes such as *tri-* or *octa-* denote peptides with three or eight *residues*, respectively. By convention, peptides are written with the residue that bears the free  $-\text{NH}_2$  group on the left. To draw a peptide, use a zigzag to represent the main chain or backbone. Add the main chain atoms, which occur in the repeating order:  $-\text{N}-$ ,  $-\text{C}(=\text{O})-$ ,  $-\text{N}-$ ,  $-\text{C}(=\text{O})-$ . Now add a hydrogen atom to each  $-\text{C}(=\text{O})-$  carbon and to each peptide nitrogen, and an oxygen atom to the carbonyl carbon. Finally, add the appropriate R groups (shaded) to each  $-\text{C}(=\text{O})-$  carbon atom.



Three-letter abbreviations linked by straight lines represent an unambiguous primary structure. Lines are omitted for single-letter abbreviations.

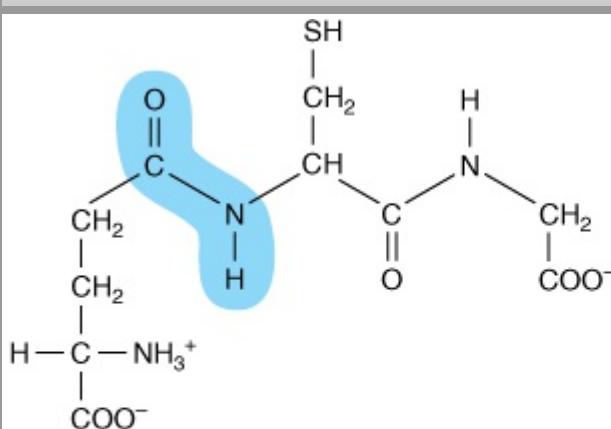
Glu - Ala - Lys - Gly - Tyr - Ala

E   A   K   G   Y   A

### Some Peptides Contain Unusual Amino Acids

In mammals, peptide hormones typically contain only the 20 genetically encoded  $\alpha$ -amino acids linked by standard peptide bonds. Other peptides may, however, contain nonprotein amino acids, derivatives of the protein amino acids, or amino acids linked by an atypical peptide bond. For example, the amino terminal glutamate of glutathione, a tripeptide that participates in protein folding and in the metabolism of xenobiotics (Chapter 53), is linked to cysteine by a non- $\alpha$ - peptide bond (Figure 3-3). The amino terminal glutamate of thyrotropin-releasing hormone (TRH) is cyclized to pyroglutamic acid, and the carboxyl group of the carboxyl terminal prolyl residue is amidated. The nonprotein amino acids D-phenylalanine and ornithine are present in the cyclic peptide antibiotics tyrocidin and gramicidin S, while the heptapeptide opioids dermorphin and deltorphin in the skin of South American tree frogs contain D-tyrosine and D-alanine.

**Figure 3-3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

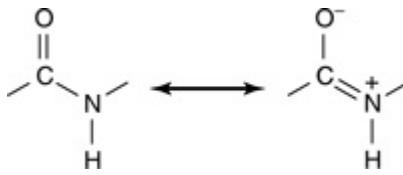
**Glutathione ( $\gamma$ -glutamyl-cysteinyl-glycine).** Note the non- $\alpha$ - peptide bond that links Glu to Cys.

### Peptides Are Polyelectrolytes

The peptide bond is uncharged at any pH of physiologic interest. Formation of peptides from amino acids is therefore accompanied by a net loss of one positive and one negative charge per peptide bond formed. Peptides nevertheless are charged at physiologic pH owing to their terminal carboxyl and amino groups and, where present, their acidic or basic R groups. As for amino acids, the net charge on a peptide depends on the pH of its environment and on the  $pK_a$  values of its dissociating groups.

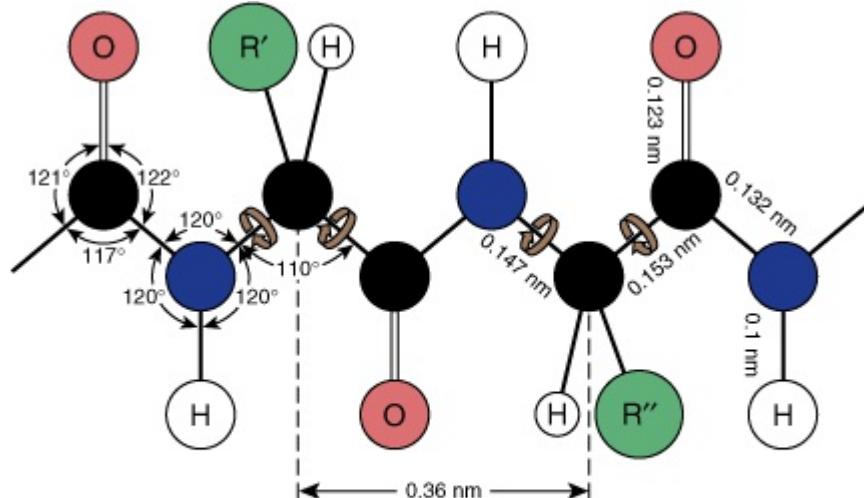
### The Peptide Bond Has a Partial Double-Bond Character

Although peptides are written as if a single bond linked the  $\alpha$ -carboxyl and  $\alpha$ -nitrogen atoms, this bond in fact exhibits a partial double-bond character:



**There thus is no freedom of rotation about the bond that connects the carbonyl carbon and the nitrogen of a peptide bond.** Consequently, the O, C, N, and H atoms of a peptide bond are coplanar. The imposed semirigidity of the peptide bond has important consequences for the manner in which peptides and proteins fold to generate higher order of structure. Encircling brown arrows (Figure 3–4) indicate free rotation about the remaining bonds of the polypeptide backbone.

**Figure 3-4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Dimensions of a fully extended polypeptide chain.** The four atoms of the peptide bond are coplanar. Free rotation can occur about the bonds that connect the  $\alpha$ -carbon with the  $\alpha$ -nitrogen and with the  $\alpha$ -carboxyl carbon (brown arrows). The extended polypeptide chain is thus a semirigid structure with two-thirds of the atoms of the backbone held in a fixed planar relationship one to another. The distance between adjacent  $\alpha$ -carbon atoms is 0.36 nm (3.6 Å). The interatomic distances and bond angles, which are not equivalent, are also shown. (Redrawn and reproduced, with permission, from Pauling L, Corey LP, Branson HR: The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci USA 1951;37:205.)

## Noncovalent Forces Constrain Peptide Conformations

Folding of a peptide probably occurs coincident with its biosynthesis (see Chapter 37). The physiologically active conformation reflects the collective contributions of the amino acid sequence, steric hindrance, and noncovalent interactions (eg, hydrogen bonding, hydrophobic interactions) between residues. Common conformations include  $\alpha$ -helices and  $\beta$ -pleated sheets (see Chapter 5).

## ANALYSIS OF THE AMINO ACID CONTENT OF BIOLOGIC MATERIALS

To determine the identity of each amino acid present in a protein, it is first treated with hot hydrochloric acid to hydrolyze the peptide bonds. There are several methods for separation and identification of amino acids derived from a protein hydrolysate or from urine or other biologic fluids. One approach is to react the amino acids with 6-amino-*N*-hydroxysuccinimidyl carbamate to form fluorescent derivatives that can be separated by high-pressure liquid chromatography (see Chapter 4). An alternative approach, which requires only minimal equipment, employs partition chromatography on a solid support, typically a sheet of filter paper (paper chromatography) or a thin layer of powdered cellulose or silica gel on an inert support (thin-layer chromatography, or TLC). The amino acids present are resolved by a mobile phase that contains a mixture of miscible polar and nonpolar components (eg, *n*-butanol, formic acid, and water). As the mobile phase moves up the sheet, its polar components associate with the polar groups of the support. The solvent therefore becomes progressively less polar as it migrates up the sheet. The amino acids therefore partition between a polar stationary phase and a less polar mobile phase ("partition chromatography"). Nonpolar amino acids (eg, Leu, Ile) migrate the farthest as they spend the greatest proportion of their time in the mobile phase. Polar amino acids (eg, Glu, Lys) travel the least distance from the origin as they spend a high proportion of their time in the stationary phase consisting of a layer of polar solvent molecules immobilized by their association with the cellulose or silica support. Following removal of the solvent by air drying, amino acids are visualized using ninhydrin, which forms purple products with  $\alpha$ -amino acids, but a yellow adduct with proline and hydroxyproline.

## SUMMARY

- Both D-amino acids and non-L-amino acids occur in nature, but only L-amino acids are present in proteins.
- All amino acids possess at least two weakly acidic functional groups, R-NH<sub>3</sub><sup>+</sup> and R-COOH. Many also possess additional weakly acidic functional groups such as —OH, —SH, guanidino, or imidazole moieties.
- The pK<sub>a</sub> values of all functional groups of an amino acid dictate its net charge at a given pH. pI is the pH at which an amino acid bears no net charge and thus does not move in a direct current electrical field.
- Of the biochemical reactions of amino acids, the most important is the formation of peptide bonds.
- The R groups of amino acids determine their unique biochemical functions. Amino acids are classified as basic, acidic, aromatic, aliphatic, or sulfur containing based on the properties of their R groups.
- Peptides are named for the number of amino acid residues present, and as derivatives of the carboxyl terminal residue. The primary structure of a peptide is its amino acid sequence, starting from the amino-terminal residue.
- The partial double-bond character of the bond that links the carbonyl carbon and the nitrogen of a peptide renders four atoms of the peptide bond coplanar and restricts the number of possible peptide conformations.

## REFERENCES

- Doolittle RF: Reconstructing history with amino acid sequences. *Protein Sci* 1992;1:191.
- Gladyshev VN, Hatfield DL: Selenocysteine-containing proteins in mammals. *J Biomed Sci* 1999;6:151.
- Kolodkin-Gal I: D-Amino acids trigger biofilm disassembly. *Science* 2010;328:627.
- Kreil G: D-Amino acids in animal peptides. *Annu Rev Biochem* 1997;66:337.
- Nokihara K, Gerhardt J: Development of an improved automated gas-chromatographic chiral analysis system: application to nonnatural amino acids and natural protein hydrolysates. *Chirality* 2001;13:431.
- Papp LV: From selenium to selenoproteins: synthesis, identity, and their role in human health. *Antioxidants Redox Signal* 2007;9:775.
- Sanger F: Sequences, sequences, and sequences. *Annu Rev Biochem* 1988;57:1.
- Stadtman TC: Selenocysteine. *Annu Rev Biochem* 1996;65:83.
- Wilson NA, Barbar E, Fuchs JA, et al: Aspartic acid 26 in reduced *Escherichia coli* thioredoxin has a pKa greater than 9. *Biochemistry* 1995;34:8931.

## OBJECTIVES

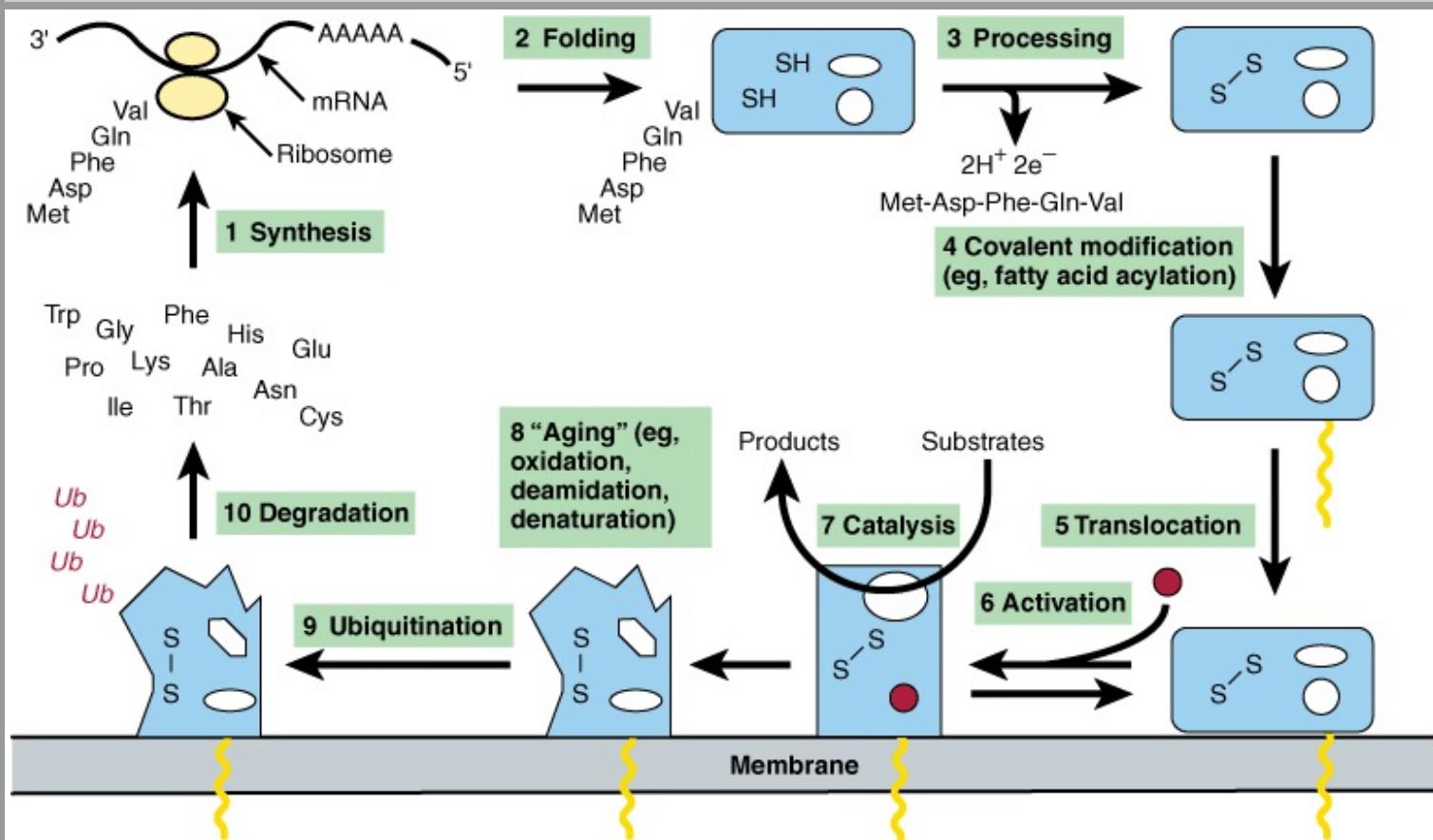
After studying this chapter, you should be able to:

- Describe multiple chromatographic methods commonly employed for the isolation of proteins from biologic materials.
- Explain how scientists analyze the sequence or structure of a protein to extract insights into its possible physiologic function.
- List several of the posttranslational alterations that proteins undergo during their lifetime and the influence of such modifications upon a protein's function and fate.
- Describe the chemical basis of the Edman method for determining primary structure.
- Give three reasons why mass spectrometry (MS) has largely supplanted chemical methods for the determination of the primary structure of proteins and the detection of posttranslational modifications.
- Explain why MS can detect posttranslational modifications that are not detected by Edman sequencing or DNA sequencing.
- Describe how DNA cloning and molecular biology made the determination of the primary structures of proteins much more rapid and efficient.
- Explain what is meant by "the proteome" and cite examples of its ultimate potential significance.
- Comment on the contributions of genomics, computer algorithms, and databases to the identification of the open reading frames (ORFs) that encode a given protein.

## BIOMEDICAL IMPORTANCE

Proteins are physically and functionally complex macromolecules that perform multiple critically important roles. For example, an internal protein network, the cytoskeleton (Chapter 49) maintains cellular shape and physical integrity. Actin and myosin filaments form the contractile machinery of muscle (Chapter 49). Hemoglobin transports oxygen (Chapter 6), while circulating antibodies defend against foreign invaders (Chapter 50). Enzymes catalyze reactions that generate energy, synthesize and degrade biomolecules, replicate and transcribe genes, process mRNAs, etc (Chapter 7). Receptors enable cells to sense and respond to hormones and other environmental cues (Chapters 41 and 42). Proteins are subject to physical and functional changes that mirror the life cycle of the organisms in which they reside. A typical protein is "born" at translation (Chapter 37), matures through posttranslational processing events such as selective proteolysis (Chapters 9 and 37), alternates between working and resting states through the intervention of regulatory factors (Chapter 9), ages through oxidation/deamidation, etc (Chapter 52), and "dies" when degraded to its component amino acids (Chapter 29). An important goal of molecular medicine is to identify biomarkers such as proteins and/or modifications to proteins whose presence, absence, or deficiency is associated with specific physiologic states or diseases (**Figure 4-1**).

**Figure 4-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Diagrammatic representation of the life cycle of a hypothetical protein.** (1) The life cycle begins with the synthesis on a ribosome of a polypeptide chain, whose primary structure is dictated by an mRNA. (2) As synthesis proceeds, the polypeptide begins to fold into its native conformation (blue). (3) Folding may be accompanied by processing events such as proteolytic cleavage of an N-terminal leader sequence (Met-Asp-Phe-Gln-Val) or the formation of disulfide bonds (S-S). (4) Subsequent covalent modifications may, for example, attach a fatty acid molecule (yellow) for (5) translocation of the modified protein to a membrane. (6) Binding an allosteric effector (red) may trigger the adoption of a catalytically active conformation. (7) Over time, proteins get damaged by chemical attack, deamidation, or denaturation, and (8) may be "labeled" by the covalent attachment of several ubiquitin molecules (Ub). (9) The ubiquitinated protein is subsequently degraded to its component amino acids, which become available for the synthesis of new proteins.

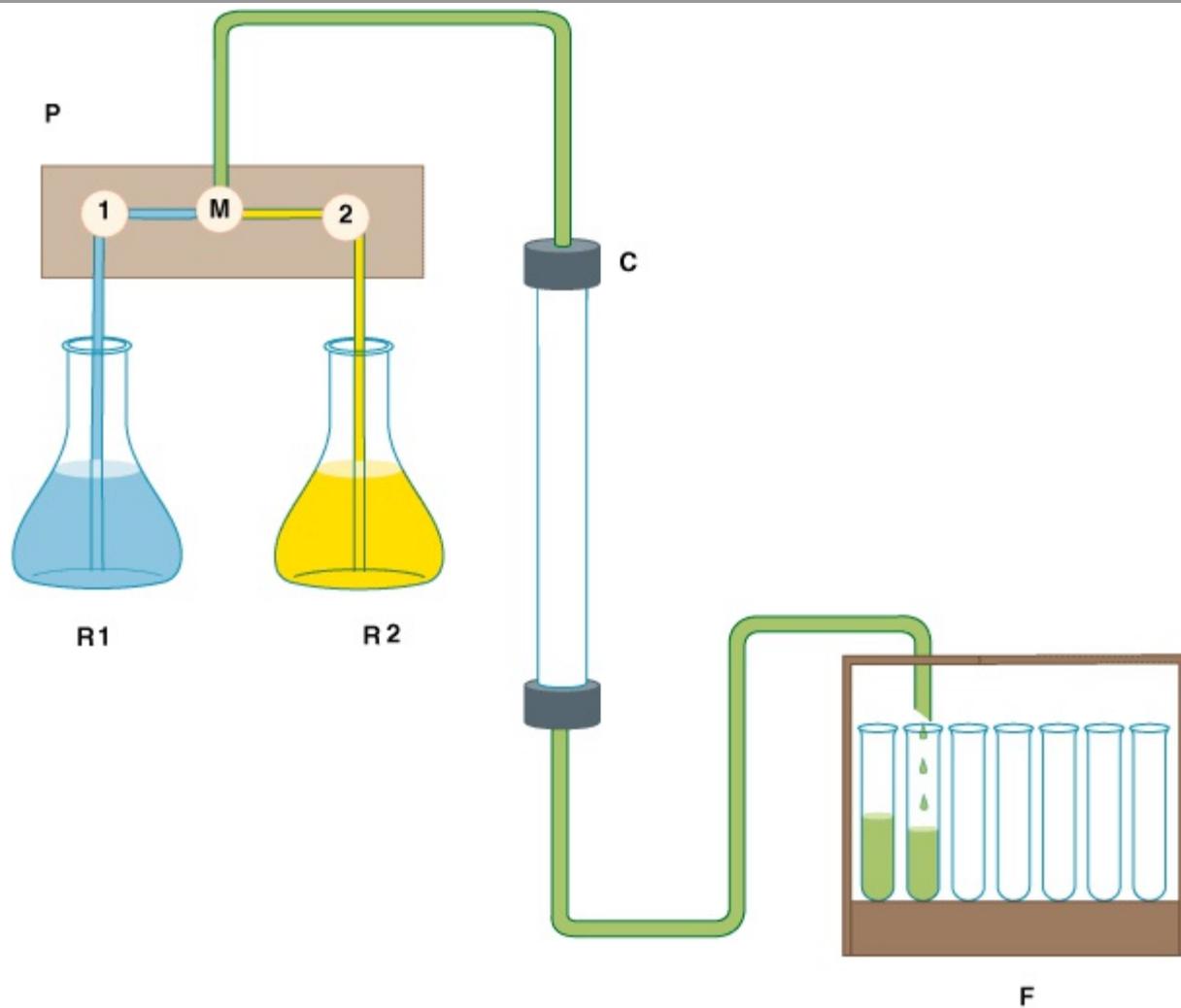
## PROTEINS & PEPTIDES MUST BE PURIFIED PRIOR TO ANALYSIS

Highly purified protein is essential for the detailed examination of its physical and functional properties. Cells contain thousands of different proteins, each in widely varying amounts. The isolation of a specific protein in quantities sufficient for analysis of its properties thus presents a formidable challenge that may require successive application of multiple purification techniques. Selective precipitation exploits differences in relative solubility of individual proteins as a function of pH (isoelectric precipitation), polarity (precipitation with ethanol or acetone), or salt concentration (salting out with ammonium sulfate). Chromatographic techniques separate one protein from another based upon difference in their size (size exclusion chromatography), charge (ion-exchange chromatography), hydrophobicity (hydrophobic interaction chromatography), or ability to bind a specific ligand (affinity chromatography).

### Column Chromatography

In column chromatography, the stationary phase matrix consists of small beads loaded into a cylindrical container of glass, plastic, or steel called a column. Liquid-permeable frits confine the beads within this space while allowing the mobile-phase liquid to flow or percolate through the column. The stationary phase beads can be chemically derivatized to coat their surface with the acidic, basic, hydrophobic, or ligand-like groups required for ion exchange, hydrophobic interaction, or affinity chromatography. As the mobile-phase liquid emerges from the column, it is automatically collected in a series of small portions called fractions. **Figure 4-2** depicts the basic arrangement of a simple bench-top chromatography system.

**Figure 4-2**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Components of a typical liquid chromatography apparatus.** R1 and R2: Reservoirs of mobile-phase liquid. P: Programmable pumping system containing two pumps, 1 and 2, and a mixing chamber, M. The system can be set to pump liquid from only one reservoir, to switch reservoirs at some predetermined point to generate a step gradient, or to mix liquids from the two reservoirs in proportions that vary over time to create a continuous gradient. C: Glass, metal, or plastic column containing stationary phase. F: Fraction collector for collecting portions, called *fractions*, of the eluant liquid in separate test tubes.

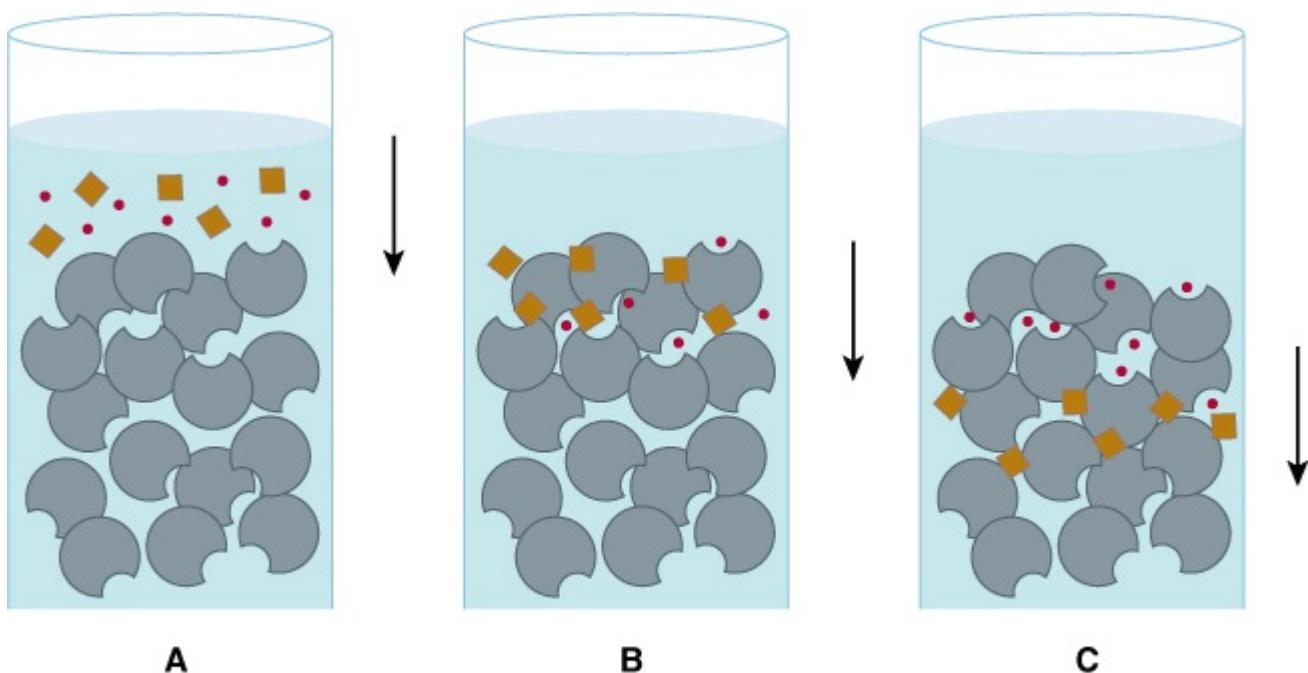
### HPLC—High-Pressure Liquid Chromatography

First-generation column chromatography matrices consisted of long, intertwined oligosaccharide polymers shaped into spherical beads roughly a tenth of a millimeter in diameter. Unfortunately, their relatively large size perturbed mobile-phase flow and limited the available surface area. Reducing particle size offered the potential to greatly increase resolution. However, the resistance created by the more tightly packed matrix required the use of very high pressures that would crush the soft and spongy polysaccharide beads and similar materials, eg, acrylamide. Eventually, methods were developed to manufacture silicon particles of the necessary size and shape, to derivatize their surface with various functional groups, and to pack them into stainless steel columns capable of withstanding pressures of several thousand psi. Because of their greater resolving power, high-pressure liquid chromatography systems have largely displaced the once familiar glass columns in the protein purification laboratory.

### Size-Exclusion Chromatography

Size exclusion—or gel filtration—chromatography separates proteins based on their **Stokes radius**; the radius of the sphere they occupy as they tumble in solution. The Stokes radius is a function of molecular mass and shape. A tumbling elongated protein occupies a larger volume than a spherical protein of the same mass. Size-exclusion chromatography employs porous beads (Figure 4-3). The pores are analogous to indentations in a riverbank. As objects move downstream, those that enter an indentation are retarded until they drift back into the main current. Similarly, proteins with Stokes radii too large to enter the pores (excluded proteins), remain in the flowing mobile phase, and emerge *before* proteins that can enter the pores (included proteins). Proteins thus emerge from a gel filtration column in descending order of their Stokes radii.

**Figure 4-3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Size-exclusion chromatography.** A: A mixture of large molecules (brown) and small molecules (red) are applied to the top of a gel filtration column. B: Upon entering the column, the small molecules enter pores in the stationary phase matrix (gray) from which the large molecules are excluded. C: As the mobile phase (blue) flows down the column, the large, excluded molecules flow with it, while the small molecules, which are temporarily sheltered from the flow when inside the pores, lag farther and farther behind.

## Ion-Exchange Chromatography

In ion-exchange chromatography, proteins interact with the stationary phase by charge-charge interactions. Proteins with a net positive charge at a given pH will tightly adhere to beads with negatively charged functional groups such as carboxylates or sulfates (cation exchangers). Similarly, proteins with a net negative charge adhere to beads with positively charged functional groups, typically tertiary, or quaternary amines (anion exchangers). Nonadherent proteins flow through the matrix and are washed away. Bound proteins are then selectively displaced by gradually raising the ionic strength of the mobile phase, thereby weakening charge-charge interactions. Proteins elute in inverse order of the strength of their interactions with the stationary phase.

## Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography separates proteins based on their tendency to associate with a stationary phase matrix coated with hydrophobic groups (eg, phenyl Sephadex). Proteins with exposed hydrophobic surfaces adhere to the matrix via hydrophobic interactions that are enhanced by employing a mobile phase of high ionic strength. After nonadherent proteins are washed away, the polarity of the mobile phase is decreased by gradually lowering the salt concentration of the flowing mobile phase. If the interaction between protein and stationary phase is particularly strong, ethanol or glycerol may be added to the mobile phase to decrease its polarity and further weaken hydrophobic interactions.

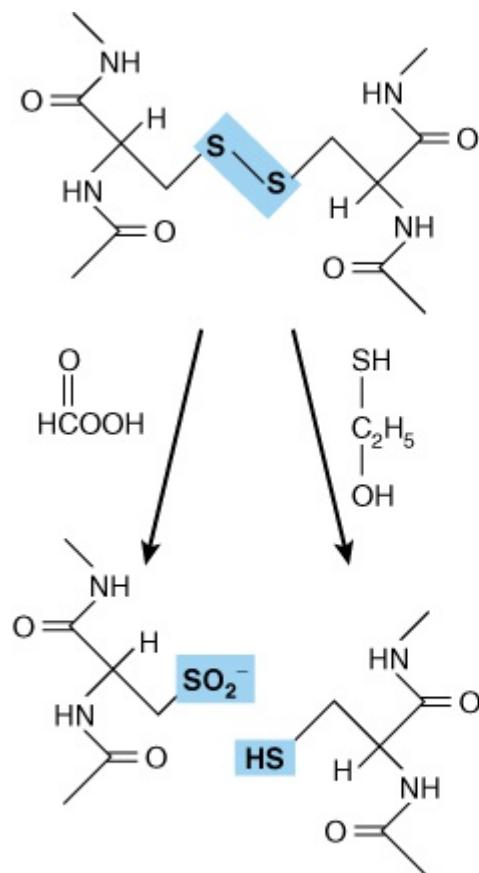
## Affinity Chromatography

Affinity chromatography exploits the high selectivity of most proteins for their ligands. Enzymes may be purified by affinity chromatography using immobilized substrates, products, coenzymes, or inhibitors. In theory, only proteins that interact with the immobilized ligand adhere. Bound proteins are then eluted either by competition with free, soluble ligand or, less selectively, by disrupting protein-ligand interactions using urea, guanidine hydrochloride, mildly acidic pH, or high salt concentrations. Commercially available stationary phase matrices contain ligands such as NAD<sup>+</sup> or ATP analogs. Purification of recombinantly expressed proteins is frequently facilitated by modifying the cloned gene to add a new fusion domain designed to interact with a specific matrix-bound ligand (Chapter 7).

## Protein Purity Is Assessed by Polyacrylamide Gel Electrophoresis (PAGE)

The most widely used method for determining the purity of a protein is SDS-PAGE—polyacrylamide gel electrophoresis (PAGE) in the presence of the anionic detergent sodium dodecyl sulfate (SDS). Electrophoresis separates charged biomolecules based on the rates at which they migrate in an applied electrical field. For SDS-PAGE, acrylamide is polymerized and cross-linked to form a porous matrix. SDS binds to proteins at a ratio of one molecule of SDS per two peptide bonds, causing the polypeptide to unfold or denature. When used in conjunction with 2-mercaptoethanol or dithiothreitol to reduce and break disulfide bonds (Figure 4–4), SDS-PAGE separates the component polypeptides of multimeric proteins. The large number of anionic SDS molecules, each bearing a charge of  $-1$ , overwhelms the charge contributions of the amino acid functional groups endogenous to the polypeptides. Since the charge-to-mass ratio of each SDS-polypeptide complex is approximately equal, the physical resistance each peptide encounters as it moves through the acrylamide matrix determines the rate of migration. Since large complexes encounter greater resistance, polypeptides separate based on their relative molecular mass ( $M_r$ ). Individual polypeptides trapped in the acrylamide gel after removal of the electrical field are visualized by staining with dyes such as Coomassie blue (Figure 4–5).

**Figure 4-4**

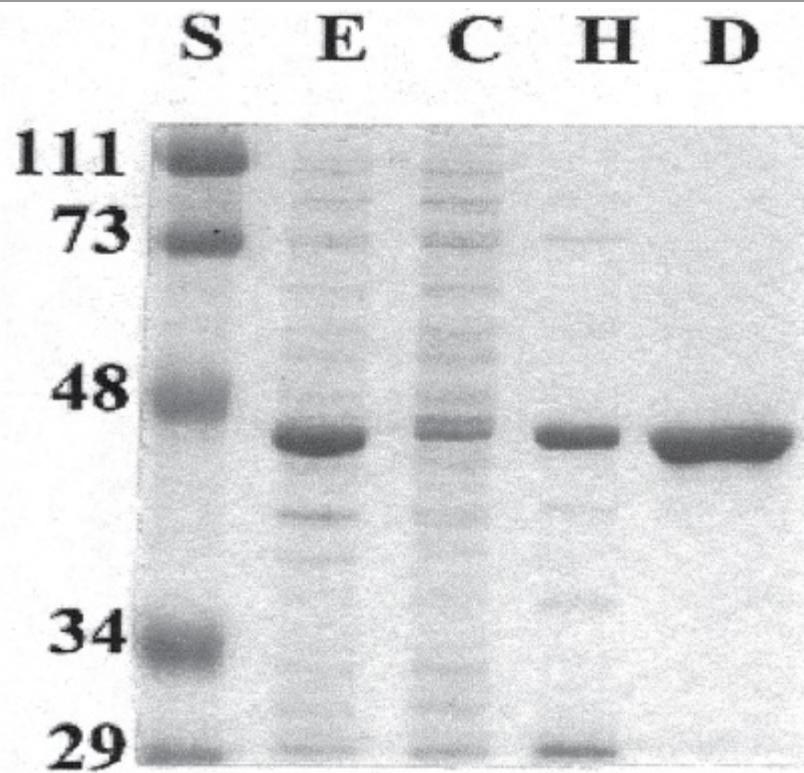


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oxidative cleavage of adjacent polypeptide chains linked by disulfide bonds (highlighted in blue) by performic acid (left) or reductive cleavage by mercaptoethanol (right) forms two peptides that contain cysteic acid residues or cysteinyl residues, respectively.

**Figure 4-5**



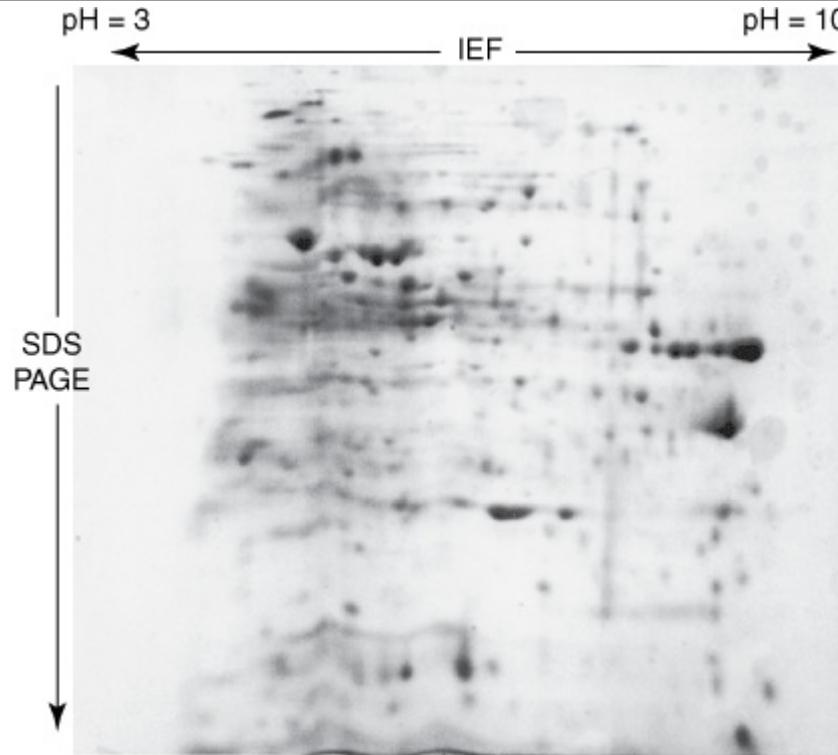
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Use of SDS-PAGE to observe successive purification of a recombinant protein.** The gel was stained with Coomassie blue. Shown are protein standards (lane S) of the indicated  $M_r$ , in kDa, crude cell extract (E), cytosol (C), high-speed supernatant liquid (H), and the DEAE-Sepharose fraction (D). The recombinant protein has a mass of about 45 kDa.

Ionic buffers called ampholytes and an applied electric field are used to generate a pH gradient within a polyacrylamide matrix. Applied proteins migrate until they reach the region of the matrix where the pH matches their isoelectric point ( $pI$ ), the pH at which a molecule's net charge is 0. IEF is used in conjunction with SDS-PAGE for two-dimensional electrophoresis, which separates polypeptides based on  $pI$  in one dimension and on  $M_r$  in the second (Figure 4–6). Two-dimensional electrophoresis is particularly well suited for separating the components of complex mixtures of proteins.

**Figure 4-6**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Two-dimensional IEF-SDS-PAGE.** The gel was stained with Coomassie blue. A crude bacterial extract was first subjected to isoelectric focusing (IEF) in a pH 3–10 gradient. The IEF gel was then placed horizontally on the top of an SDS-PAGE gel, and the proteins then further resolved by SDS-PAGE. Notice the greatly improved resolution of distinct polypeptides relative to ordinary SDS-PAGE gel (Figure 4–5).

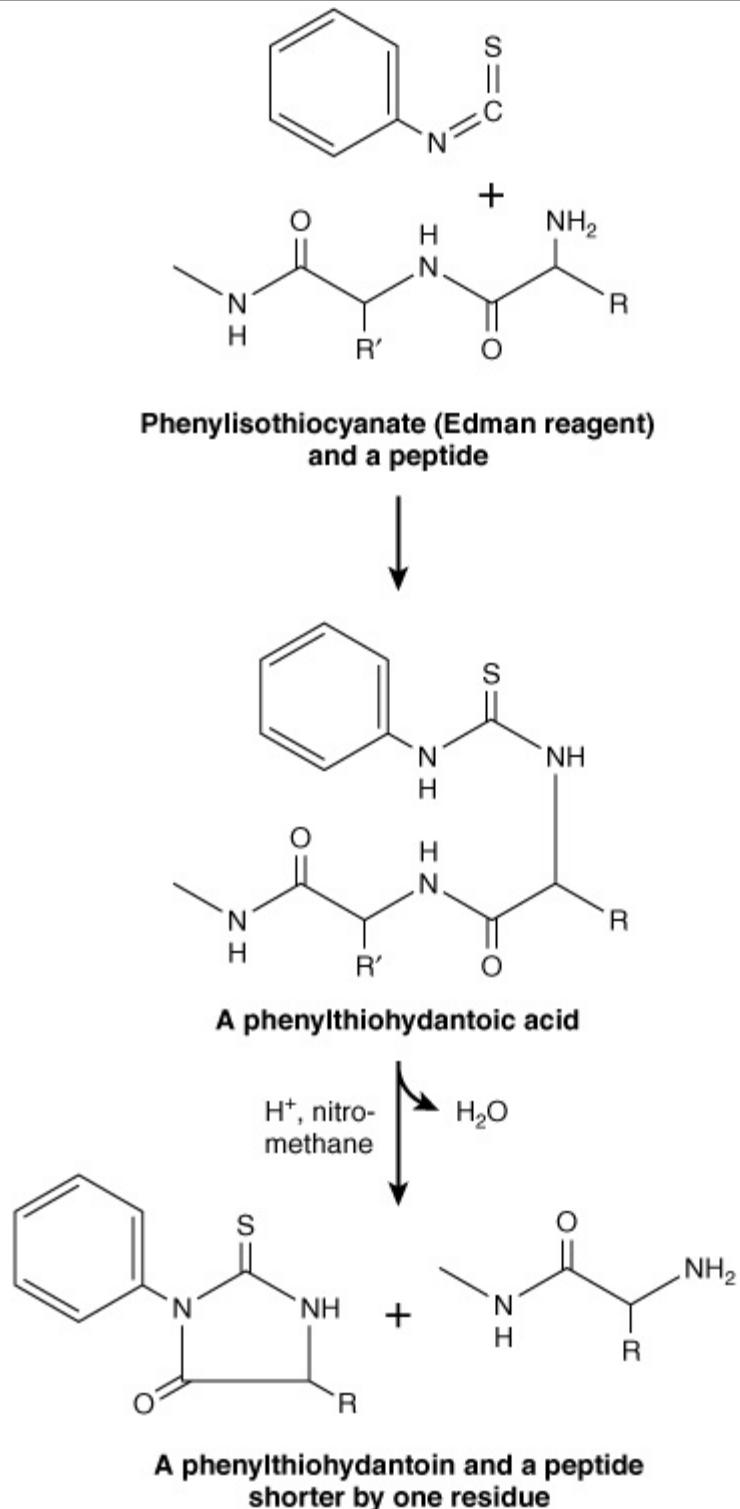
## SANGER WAS THE FIRST TO DETERMINE THE SEQUENCE OF A POLYPEPTIDE

Mature insulin consists of the 21-residue A chain and the 30-residue B chain linked by disulfide bonds. Frederick Sanger reduced the disulfide bonds (Figure 4-4), separated the A and B chains, and cleaved each chain into smaller peptides using trypsin, chymotrypsin, and pepsin. The resulting peptides were then isolated and treated with acid to hydrolyze a portion of the peptide bonds and generate peptides with as few as two or three amino acids. Each peptide was reacted with 1-fluoro-2,4-dinitrobenzene (Sanger's reagent), which derivatizes the exposed  $\alpha$ -amino groups of the amino-terminal residues. The amino acid content of each peptide was then determined and the amino-terminal amino acid identified. The  $\alpha$ -amino group of lysine also reacts with Sanger's reagent; but since an amino-terminal lysine reacts with 2 mol of Sanger's reagent, it is readily distinguished from a lysine in the interior of a peptide. Working from di- and tripeptides up through progressively larger fragments, Sanger was able to reconstruct the complete sequence of insulin, an accomplishment for which he received a Nobel Prize, in 1958.

## THE EDMAN REACTION ENABLES PEPTIDES & PROTEINS TO BE SEQUENCED

Pehr Edman introduced phenylisothiocyanate (Edman's reagent) to selectively label the amino-terminal residue of a peptide. In contrast to Sanger's reagent, the phenylthiohydantoin (PTH) derivative can be removed under mild conditions to generate a new amino-terminal residue (Figure 4-7). Successive rounds of derivatization with Edman's reagent can therefore be used to sequence many residues of a single sample of peptide. Even so, the determination of the complete sequence of a protein by chemical methods remains a time- and labor-intensive process to this day.

**Figure 4-7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**The Edman reaction.** Phenylisothiocyanate derivatizes the amino-terminal residue of a peptide as a phenylthiohydantoic acid. Treatment with acid in a nonhydroxylic solvent releases a phenylthiohydantoin, which is subsequently identified by its chromatographic mobility, and a peptide one residue shorter. The process is then repeated.

The heterogeneous chemical properties of the amino acids meant that every step in the procedure represented a compromise between efficiency for any particular amino acid or set of amino acids and the flexibility needed to accommodate all 20. Consequently, each step in the process operates at less than 100% efficiency, which leads to the accumulation of polypeptide fragments with varying *N*-termini. Eventually, it becomes impossible to distinguish the correct PTH amino acid for that position in the peptide from the contaminants. As a result, the read length for Edman sequencing varies from 5 to 30 amino acid residues depending upon the quantity and purity of the peptide.

In order to determine the complete sequence of a polypeptide several hundred residues in length, a protein must first be cleaved into smaller peptides, using either a protease or a reagent such as cyanogen bromide. Following purification by reversed phase high-pressure liquid chromatography (HPLC), these peptides are then analyzed by Edman sequencing. In order to assemble these short peptide sequences to solve the complete sequence of the intact polypeptide, it is necessary to analyze peptides whose sequences overlap one another. This is accomplished by generating multiple sets of peptides using more than one method of cleavage. The large quantities of purified protein required to test multiple protein

fragmentation and peptide purification conditions constitutes the second major drawback of direct chemical protein sequencing techniques.

## MOLECULAR BIOLOGY REVOLUTIONIZED THE DETERMINATION OF PRIMARY STRUCTURE

The reactions that sequentially derivatize and cleave PTH amino acids from the amino-terminal end of a peptide typically are conducted in an automated sequenator. DNA sequencing, by contrast, is both far more rapid and more economical. Recombinant techniques permit researchers to manufacture a virtually infinite supply of DNA using the original sample as template (Chapter 39). DNA sequencing methods, whose chemistry was also developed by Sanger, routinely enable polydeoxyribonucleotide sequences a few hundred residues in length to be determined in single analysis, while automated sequencers can "read" sequences several thousand nucleotides in length. Knowledge of the genetic code enables the sequence of the encoded polypeptide to be determined by simply translating the oligonucleotide sequence of its gene. Conversely, early molecular biologists designed complementary oligonucleotide probes to identify the DNA clone containing the gene of interest by reversing this process and using a segment of chemically determined amino acid sequence as template. The advent of DNA cloning thus ushered in the widespread use of a hybrid approach in which Edman chemistry was employed to sequence a small portion of the protein, then exploiting this information to determine the remaining sequence by DNA cloning and sequencing.

## GENOMICS ENABLES PROTEINS TO BE IDENTIFIED FROM SMALL AMOUNTS OF SEQUENCE DATA

Today the number of organisms for which the complete DNA sequence of their genomes has been determined and made available to the scientific community numbers in the hundreds (see Chapter 10). These sequences encompass nearly all of the "model organisms" commonly employed in biomedical research laboratories: *Homo sapiens*, mouse, rat, *Escherichia coli*, *Drosophila melanogaster*, *Caenorhabditis elegans*, yeast, etc., as well as numerous pathogens. Meanwhile, across the globe, arrays of automated DNA sequencers continue to generate genome sequence data ever more rapidly and economically. Thus, for most research scientists the sequence of the protein(s) with which they are working has already been determined and lies waiting to be accessed in a database such as GenBank (Chapter 10). All that the scientist needs is to acquire sufficient amino acid sequence information from the protein, sometimes as little as five or six consecutive residues, to make an unambiguous identification. While the requisite amino acid sequence information can be obtained using the Edman technique, today mass spectrometry (MS) has emerged as the method of choice for protein identification.

## MASS SPECTROMETRY CAN DETECT COVALENT MODIFICATIONS

The superior sensitivity, speed, and versatility of MS have replaced the Edman technique as the principal method for determining the sequences of peptides and proteins. MS is significantly more sensitive and tolerant of variations in sample quality. Moreover, since mass and charge are common properties of a wide range of biomolecules, MS can be used to analyze metabolites, carbohydrates, and posttranslational modifications such as phosphorylation or hydroxylation that add readily identified increments of mass to a protein (Table 4–1). These modifications are difficult to detect using the Edman technique and undetectable in the DNA-derived amino acid sequence.

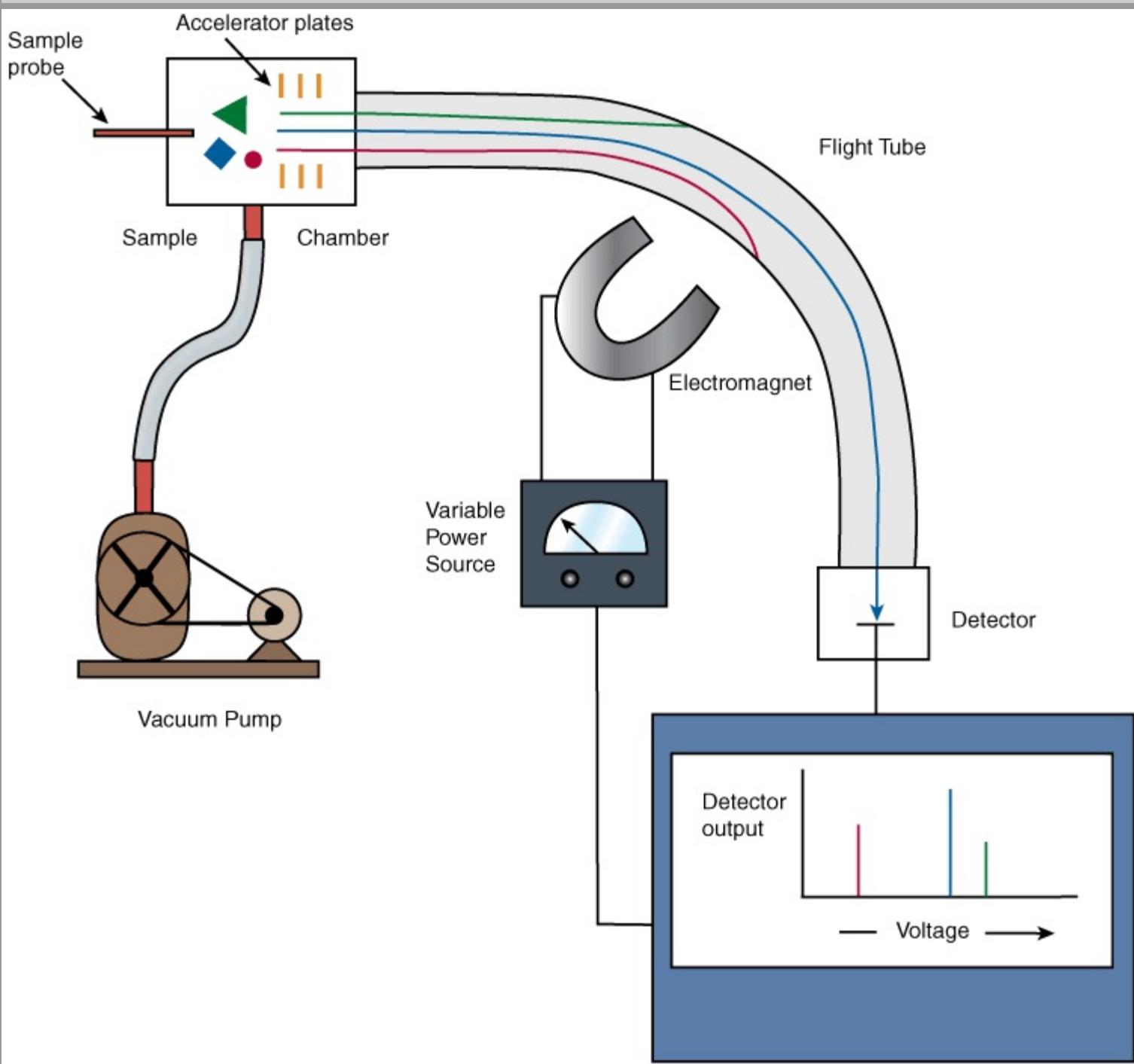
**Table 4–1 Mass Increases Resulting from Common Post-Translational Modifications**

Modification	Mass Increase (Da)
Phosphorylation	80
Hydroxylation	16
Methylation	14
Acetylation	42
Myristylation	210
Palmitoylation	238
Glycosylation	162

## MASS SPECTROMETERS COME IN VARIOUS CONFIGURATIONS

In a simple, single quadrupole mass spectrometer a sample is placed under vacuum and allowed to vaporize in the presence of a proton donor to impart a positive charge. An electrical field then propels the cations toward a curved flight tube where they encounter a magnetic field, which deflects them at a right angle to their original direction of flight (Figure 4-8). The current powering the electromagnet is gradually increased until the path of each ion is bent sufficiently to strike a detector mounted at the end of the flight tube. For ions of identical net charge, the force required to bend their path to the same extent is proportionate to their mass.

**Figure 4-8**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Basic components of a simple mass spectrometer.** A mixture of molecules, represented by a red circle, green triangle, and blue diamond, is vaporized in an ionized state in the sample chamber. These molecules are then accelerated down the flight tube by an electrical potential applied to the accelerator grid (yellow). An adjustable field strength electromagnet applies a magnetic field that deflects the flight of the individual ions until they strike the detector. The greater the mass of the ion, the higher the magnetic field required to focus it onto the detector.

The flight tube for a time-of-flight (TOF) mass spectrometer is linear. Following vaporization of the sample in the presence of a proton donor, an electric field is briefly applied to accelerate the ions toward the detector at the end of the flight tube. For molecules of identical charge, the velocity to which they are accelerated—and hence the time required to reach the detector—is inversely proportional to their mass.

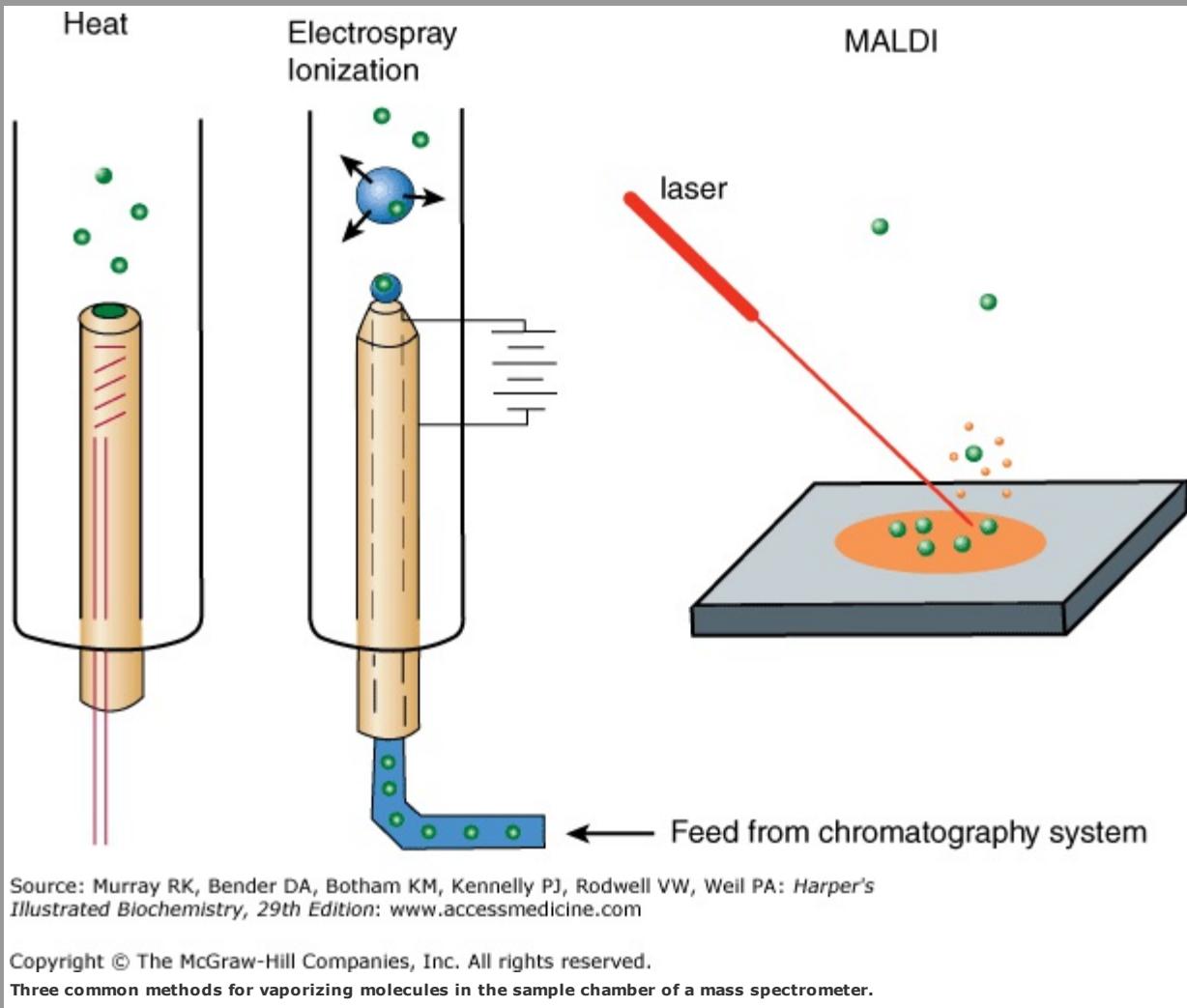
Quadrupole mass spectrometers generally are used to determine the masses of molecules of 4000 Da or less, whereas time-of-flight mass spectrometers are used to determine the large masses of complete proteins. Various combinations of multiple quadrupoles, or reflection of ions back down the linear flight tube of a TOF mass spectrometer, are used to create more sophisticated instruments.

### Peptides Can Be Volatilized for Analysis by Electrospray Ionization or Matrix-Assisted Laser Desorption

The analysis of peptides and proteins by mass spectrometry initially was hindered by difficulties in volatilizing large organic molecules. While small organic molecules could be readily vaporized by heating in a vacuum (Figure 4-9), proteins, oligonucleotides, etc., were destroyed under these conditions. Only when reliable techniques were devised for dispersing peptides, proteins, and other large biomolecules into the vapor phase was it possible to apply MS for their structural analysis and sequence determination. Dispersion into the vapor phase

is accomplished by **electrospray ionization** and **matrix-assisted laser desorption and ionization**, aka **MALDI**. In electrospray ionization, the molecules to be analyzed are dissolved in a volatile solvent and introduced into the sample chamber in a minute stream through a capillary (Figure 4–9). As the droplet of liquid emerges into the sample chamber, the solvent rapidly disperses leaving the macromolecule suspended in the gaseous phase. The charged probe serves to ionize the sample. Electrospray ionization is frequently used to analyze peptides and proteins as they elute from an HPLC or other chromatography column already dissolved in a volatile solvent. In MALDI, the sample is mixed with a liquid matrix containing a light-absorbing dye and a source of protons. In the sample chamber, the mixture is excited using a laser, causing the surrounding matrix to disperse into the vapor phase so rapidly as to avoid heating embedded peptides or proteins (Figure 4–9).

**Figure 4-9**



Peptides inside the mass spectrometer can be broken down into smaller units by collisions with neutral helium or argon atoms (collision-induced dissociation) and the masses of the individual fragments determined. Since peptide bonds are much more labile than carbon–carbon bonds, the most abundant fragments will differ from one another by units equivalent to one or two amino acids. Since—with the exceptions of (1) leucine and isoleucine and (2) glutamine and lysine—the molecular mass of each amino acid is unique, the sequence of the peptide can be reconstructed from the masses of its fragments.

### Tandem Mass Spectrometry

Complex peptide mixtures can now be analyzed, without prior purification, by tandem MS, which employs the equivalent of two mass spectrometers linked in series. For this reason, such tandem instruments are often referred to as **MS-MS**. The first mass spectrometer separates individual peptides based upon their differences in mass. By adjusting the field strength of the first magnet, a single peptide can be directed into the second mass spectrometer, where fragments are generated and their masses determined. Alternatively, they can be held in an **ion trap** placed between the two quadrupoles and selectively passed to the second quadrupoles instead of being lost when the first quadrupoles is set to select ions of a different mass.

### Tandem Mass Spectrometry Can Detect Metabolic Abnormalities

Tandem MS can be used to screen blood samples from newborns for the presence and concentrations of amino acids, fatty acids, and other metabolites. Abnormalities in metabolite levels can serve as diagnostic indicators for a variety of genetic disorders, such as phenylketonuria, ethylmalonic encephalopathy, and glutaric aciduria type 1.

## PROTEOMICS & THE PROTEOME

### The Goal of Proteomics Is to Identify the Entire Complement of Proteins Elaborated by a Cell under Diverse Conditions

While the sequence of the human genome is known, the picture provided by genomics alone is both static and incomplete. Proteomics aims to identify the entire complement of proteins elaborated by a cell under diverse conditions. As genes are switched on and off, proteins are synthesized in particular cell types at specific times of growth or differentiation and in response to external stimuli. Muscle cells express proteins not expressed by neural cells, and the type of subunits present in the hemoglobin tetramer undergo change pre- and postpartum. Many proteins undergo post-translational modifications during maturation into functionally competent forms or as a means of regulating their properties. Knowledge of the human genome therefore represents only the beginning of the task of describing living organisms in molecular detail and understanding the dynamics of processes such as growth, aging, and disease. As the human body contains thousands of cell types, each containing thousands of proteins, the **proteome**—the set of all the proteins expressed by an individual cell at a particular time—represents a moving target of formidable dimensions.

### Two-Dimensional Electrophoresis & Gene Array Chips Are Used to Survey Protein Expression

One goal of proteomics is the identification of proteins whose levels of expression correlate with medically significant events. The presumption is that proteins whose appearance or disappearance is associated with a specific physiologic condition or disease are linked, either directly or indirectly, to their root causes and mechanisms. Determination of the proteomes characteristic of each cell type requires the utmost efficiency in the isolation and identification of individual proteins. The contemporary approach utilizes robotic automation to speed sample preparation and large two-dimensional gels to resolve cellular proteins. Individual polypeptides are then extracted and analyzed by Edman sequencing or mass spectroscopy. While only about 1000 proteins can be resolved on a single gel, two-dimensional electrophoresis has a major advantage, in that it examines the proteins themselves.

An alternative approach, called multidimensional protein identification technology (MudPIT) employs successive rounds of chromatography to resolve the peptides produced from the digestion of a complex biologic sample into several simpler fractions that can be analyzed separately by MS. **Gene arrays, sometimes called DNA chips**, in which the expression of the mRNAs that encode proteins is detected, offer a complementary approach to proteomics. While changes in the expression of the mRNA encoding a protein do not necessarily reflect comparable changes in the level of the corresponding protein, gene arrays are more sensitive than two-dimensional gels, particularly with respect to low abundance proteins, and thus can examine a wider range of gene products.

### Bioinformatics Assists Identification of Protein Functions

The functions of a large proportion of the proteins encoded by the human genome are presently unknown. The development of protein arrays or chips for directly testing the potential functions of proteins on a mass scale remains in its infancy. However, recent advances in bioinformatics permit researchers to compare amino acid sequences to discover clues to potential properties, physiologic roles, and mechanisms of action of proteins. Algorithms exploit the tendency of nature to employ variations of a structural theme to perform similar functions in several proteins [eg, the Rossmann nucleotide binding fold to bind NAD(P)H, nuclear targeting sequences, and EF hands to bind  $\text{Ca}^{2+}$ ]. These domains generally are detected in the primary structure by conservation of particular amino acids at key positions. Insights into the properties and physiologic role of a newly discovered protein thus may be inferred by comparing its primary structure with that of known proteins.

## SUMMARY

- Long amino acid polymers or polypeptides constitute the basic structural unit of proteins, and the structure of a protein provides insight into how it fulfills its functions.
- Proteins undergo post-translational alterations during their lifetime that influence their function and determine their fate.
- The Edman method has largely been replaced by MS, a sensitive and versatile tool for determining primary structure, for identifying post-translational modifications, and for detecting metabolic abnormalities.
- DNA cloning and molecular biology coupled with protein chemistry provide a hybrid approach that greatly increases the speed and efficiency for determination of primary structures of proteins.
- Genomics—the analysis of the entire oligonucleotide sequence of an organism's complete genetic material—has provided further enhancements.
- Computer algorithms facilitate identification of the ORFs that encode a given protein by using partial sequences and peptide mass profiling to search sequence databases.
- Scientists are now trying to determine the primary sequence and functional role of every protein expressed in a living cell, known as its proteome.
- A major goal is the identification of proteins and of their post-translational modifications whose appearance or disappearance correlates with physiologic phenomena, aging, or specific diseases.

## REFERENCES

- Arnaud CH: Mass spec tackles proteins. *Chem Eng News* 2006;84:17.
- Austin CP: The impact of the completed human genome sequence on the development of novel therapeutics for human disease. *Annu Rev Med* 2004;55:1.
- Deutscher MP (editor): *Guide to Protein Purification*. Methods Enzymol, vol. 182, Academic Press, 1990 (Entire volume).
- Gwynne P, Heebner G: Mass spectrometry in drug discovery and development: from physics to pharma. *Science* 2006;313:1315.
- Kislinger T, Gramolini AO, MacLennan DH, et al: Multidimensional protein identification technology (MudPIT): technical overview of a profiling method optimized for the comprehensive proteomic investigation of normal and diseased heart tissue. *J Am Soc Mass Spectrom* 2005;16:1207.
- Kolialexi A, Anagnostopoulos AK, Mavrou A, et al: Application of proteomics for diagnosis of fetal aneuploidies and pregnancy complications. *J Proteomics* 2009;72:731.
- Levy PA: An overview of newborn screening. *J Dev Behav Pediatr* 2010;31:622.
- Schena M, Shalon D, Davis RW, et al: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467.
- Scopes RK: *Protein Purification. Principles and Practice*, 3rd ed. Springer, 1994.
- Semsarian C, Seidman CE: Molecular medicine in the 21st century. *Intern Med J* 2001;31:53.
- Sharon M, Robinson CV: The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* 2007;76:167.
- Shendure J, Ji H: Next-generation DNA sequencing. *Nature Biotechnol* 2008; 26:1135.
- Sikaroodi M, Galachantz Y, Baranova AI: Tumor markers: the potential of "omics" approach. *Curr Mol Med* 2010;10:249.
- Woodage T, Broder S: The human genome and comparative genomics: understanding human evolution, biology, and medicine. *J Gastroenterol* 2003;15:68.

## OBJECTIVES

After studying this chapter, you should be able to:

- Indicate the advantages and drawbacks of several approaches to classifying proteins.
- Explain and illustrate the primary, secondary, tertiary, and quaternary structure of proteins.
- Identify the major recognized types of secondary structure and explain supersecondary motifs.
- Describe the kind and relative strengths of the forces that stabilize each order of protein structure.
- Describe the information summarized by a Ramachandran plot.
- Indicate the present state of knowledge concerning the stepwise process by which proteins are thought to attain their native conformation.
- Identify the physiologic roles in protein maturation of chaperones, protein disulfide isomerase, and peptidylproline *cis-trans*-isomerase.
- Describe the principal biophysical techniques used to study tertiary and quaternary structure of proteins.
- Explain how genetic and nutritional disorders of collagen maturation illustrate the close linkage between protein structure and function.
- For the prion diseases, outline the overall events in their molecular pathology and name the life forms each affects.

## BIOMEDICAL IMPORTANCE

In nature, form follows function. In order for a newly synthesized polypeptide to mature into a biologically functional protein capable of catalyzing a metabolic reaction, powering cellular motion, or forming the macromolecular rods and cables that provide structural integrity to hair, bones, tendons, and teeth, it must fold into a specific three-dimensional arrangement, or **conformation**. In addition, during maturation **post-translational modifications** may add new chemical groups or remove transiently-needed peptide segments. Genetic or nutritional deficiencies that impede protein maturation are deleterious to health. Examples of the former include Creutzfeldt-Jakob disease, scrapie, Alzheimer's disease, and bovine spongiform encephalopathy ("mad cow disease"). Scurvy represents a nutritional deficiency that impairs protein maturation.

## CONFORMATION VERSUS CONFIGURATION

The terms configuration and conformation are often confused. **Configuration** refers to the geometric relationship between a given set of atoms, for example, those that distinguish L- from D-amino acids. Interconversion of *configurational* alternatives requires breaking (and reforming) covalent bonds. **Conformation** refers to the spatial relationship of every atom in a molecule. Interconversion between conformers occurs without covalent bond rupture, with retention of configuration, and typically via rotation about single bonds.

## PROTEINS WERE INITIALLY CLASSIFIED BY THEIR GROSS CHARACTERISTICS

Scientists initially approached structure–function relationships in proteins by separating them into classes based upon properties such as solubility, shape, or the presence of nonprotein groups. For example, the proteins that can be extracted from cells using aqueous solutions of physiologic pH and ionic strength are classified as **soluble**. Extraction of **integral membrane proteins** requires dissolution of the membrane with detergents. **Globular proteins** are compact, roughly spherical molecules that have **axial ratios** (the ratio of their shortest to longest dimensions) of not over 3. Most enzymes are globular proteins. By contrast, many structural proteins adopt highly extended conformations. These **fibrous proteins** possess axial ratios of 10 or more.

**Lipoproteins** and **glycoproteins** contain covalently bound lipid and carbohydrate, respectively. Myoglobin, hemoglobin, cytochromes, and many other **metalloproteins** contain tightly associated metal ions. While more precise classification schemes have emerged based upon similarity, or **homology**, in amino acid sequence and three-dimensional structure, many early classification terms remain in use.

## PROTEINS ARE CONSTRUCTED USING MODULAR PRINCIPLES

Proteins perform complex physical and catalytic functions by positioning specific chemical groups in a precise three-dimensional arrangement. The polypeptide scaffold containing these groups must adopt a conformation that is both functionally efficient and physically strong. At first glance, the biosynthesis of polypeptides comprised of tens of thousands of individual atoms would appear to be extremely challenging. When one considers that a typical polypeptide can adopt  $\sim 10^{50}$  distinct conformations, folding into the conformation appropriate to their biologic function would appear to be even more difficult. As described in Chapters 3 and 4, synthesis of the polypeptide backbones of proteins employs a small set of common building blocks or modules, the amino acids, joined by a common linkage, the peptide bond. Similarly, a stepwise modular pathway simplifies the folding and processing of newly synthesized polypeptides into mature proteins.

## FOUR ORDERS OF THE PROTEIN STRUCTURE

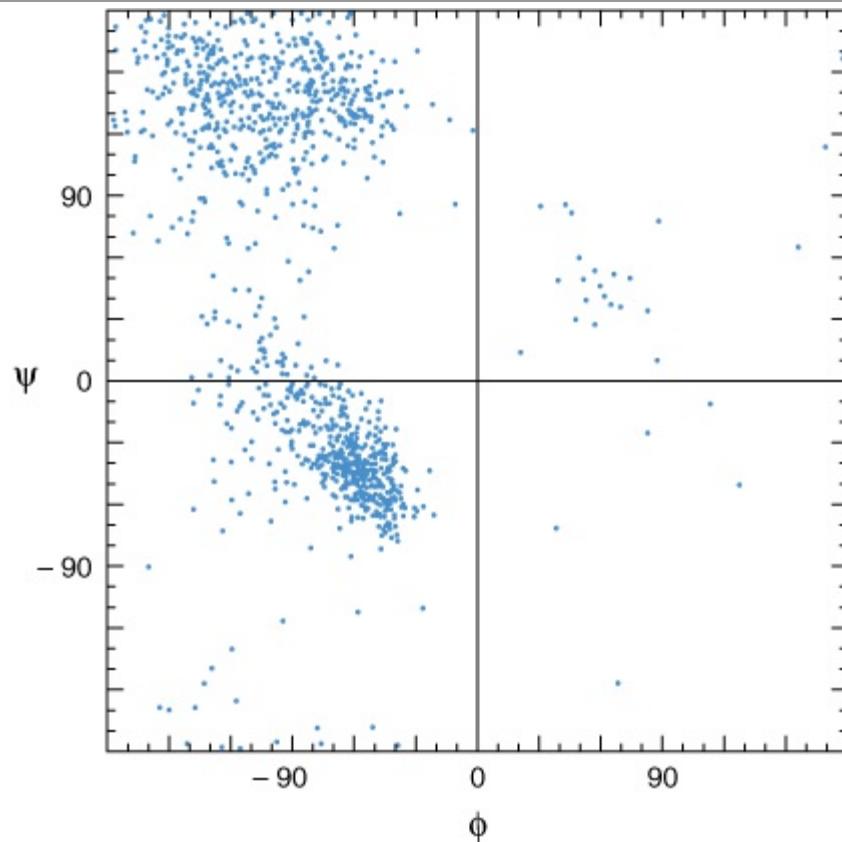
The modular nature of protein synthesis and folding are embodied in the concept of orders of the protein structure: **primary structure**—the sequence of the amino acids in a polypeptide chain; **secondary structure**—the folding of short (3- to 30-residue), contiguous segments of polypeptide into geometrically ordered units; **tertiary structure**—the assembly of secondary structural units into larger functional units such as the mature polypeptide and its component domains; and **quaternary structure**—the number and types of polypeptide units of oligomeric proteins and their spatial arrangement.

## SECONDARY STRUCTURE

### Peptide Bonds Restrict Possible Secondary Conformations

Free rotation is possible about only two of the three covalent bonds of the polypeptide backbone: the  $\text{C}=\text{O}$  bond, and the  $\text{C}-\text{N}$  bond (**Figure 3–4**). The partial double-bond character of the peptide bond that links  $\text{C}=\text{O}$  to  $\text{N}$  requires that the carbonyl carbon, carbonyl oxygen, and  $\text{N}$ -nitrogen remain coplanar thus preventing rotation. The angle about the  $\text{C}-\text{N}$  bond is termed the phi ( $\phi$ ) angle, and that about the  $\text{C}=\text{O}$  bond the psi ( $\psi$ ) angle. For amino acids other than glycine, most combinations of phi and psi angles are disallowed because of steric hindrance (**Figure 5–1**). The conformations of proline are even more restricted due to the absence of free rotation of the  $\text{N}-\text{C}=\text{O}$  bond.

**Figure 5–1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

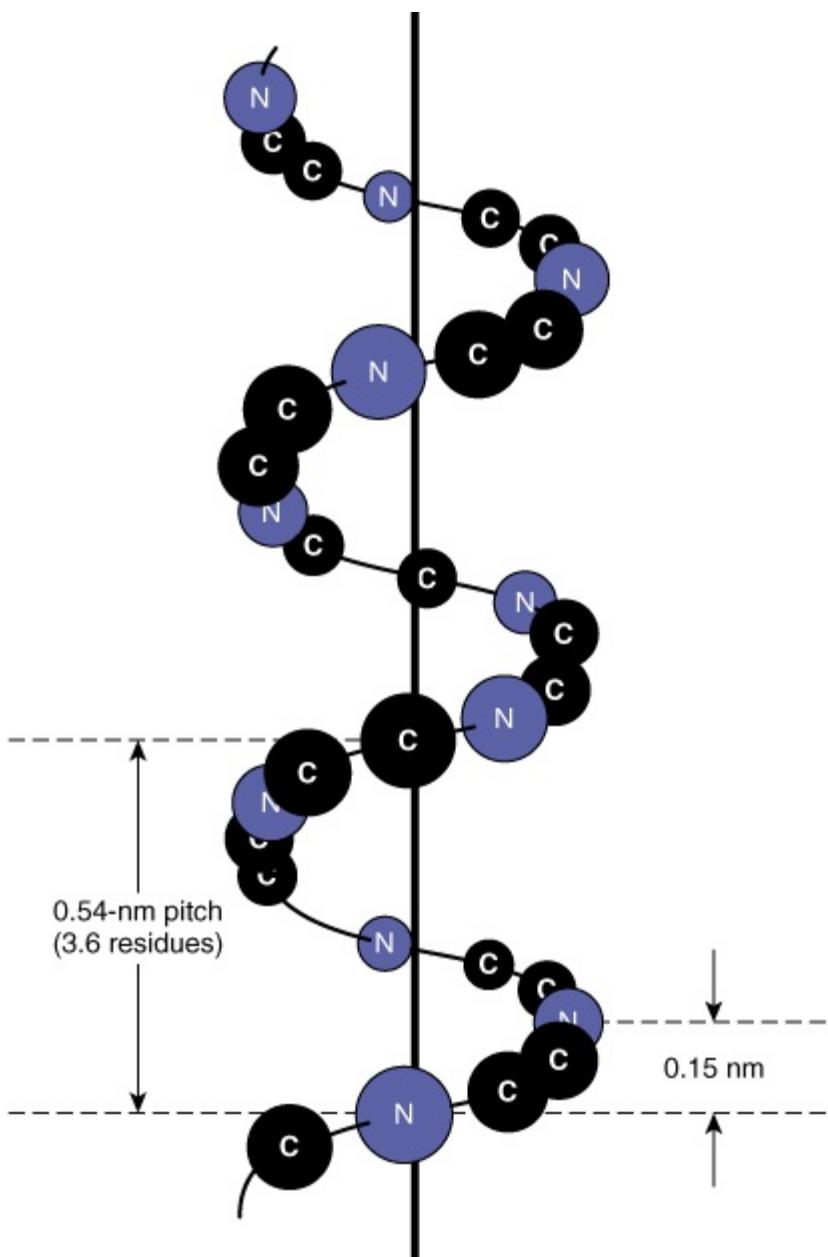
**Ramachandran plot of the main chain phi ( $\phi$ ) and psi ( $\psi$ ) angles for approximately 1000 nonglycine residues in eight proteins whose structures were solved at high resolution.** The dots represent allowable combinations, and the spaces prohibited combinations, of phi and psi angles. (Reproduced, with permission, from Richardson JS: The anatomy and taxonomy of protein structures. *Adv Protein Chem* 1981;34:167. Copyright © 1981. Reprinted with permission from Elsevier.)

Regions of ordered secondary structure arise when a series of aminoacyl residues adopt similar phi and psi angles. Extended segments of polypeptide (eg, loops) can possess a variety of such angles. The angles that define the two most common types of secondary structure, the  $\alpha$  helix and the  $\beta$  sheet, fall within the lower and upper left-hand quadrants of a Ramachandran plot, respectively (**Figure 5–1**).

### Alpha Helix

The polypeptide backbone of an  $\alpha$  helix is twisted by an equal amount about each  $\text{C}-\text{O}$  bond with a phi angle of approximately  $-57^\circ$  and a psi angle of approximately  $-47^\circ$ . A complete turn of the helix contains an average of 3.6 amino-acyl residues, and the distance it rises per turn (its *pitch*) is 0.54 nm (**Figure 5–2**). The R groups of each aminoacyl residue in an  $\alpha$  helix face outward (**Figure 5–3**). Proteins contain only L-amino acids, for which a right-handed  $\alpha$  helix is by far the more stable, and only right-handed  $\alpha$  helices are present in proteins. Schematic diagrams of proteins represent  $\alpha$  helices as coils or cylinders.

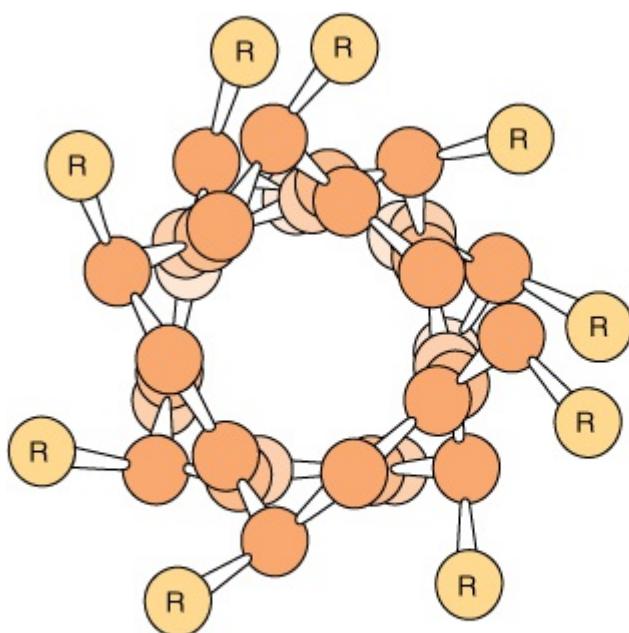
**Figure 5–2**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.  
Orientation of the main chain atoms of a peptide about the axis of an  $\alpha$  helix.

**Figure 5-3**



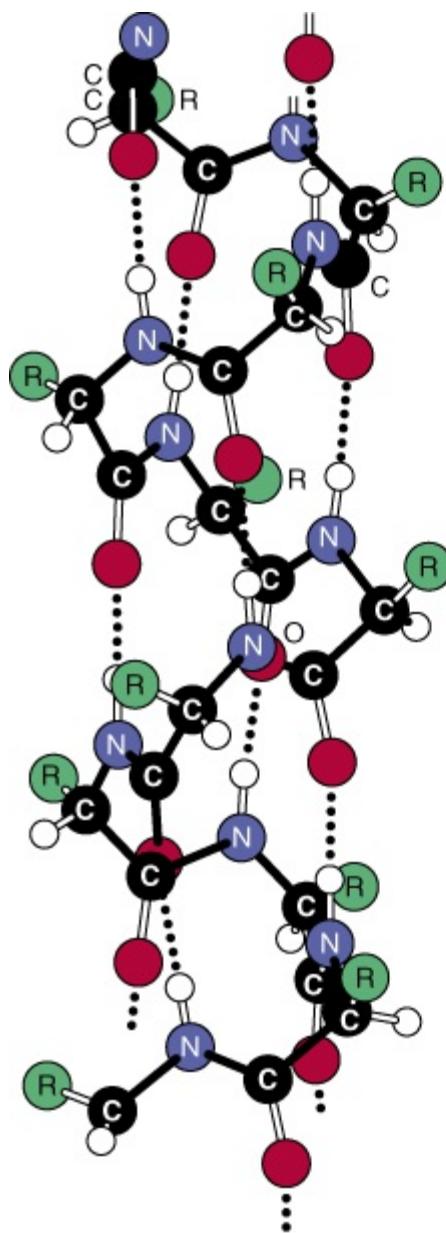
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**View down the axis of an  $\alpha$ -helix.** The side chains (R) are on the outside of the helix. The van der Waals radii of the atoms are larger than shown here; hence, there is almost no free space inside the helix. (Slightly modified and reproduced, with permission, from Stryer L: *Biochemistry*, 3rd ed. Freeman, 1995. Copyright © 1995 W.H. Freeman and Company.)

The stability of an  $\alpha$ -helix arises primarily from hydrogen bonds formed between the oxygen of the peptide bond carbonyl and the hydrogen atom of the peptide bond nitrogen of the fourth residue down the polypeptide chain (Figure 5-4). The ability to form the maximum number of hydrogen bonds, supplemented by van der Waals interactions in the core of this tightly packed structure, provides the thermodynamic driving force for the formation of an  $\alpha$ -helix. Since the peptide bond nitrogen of proline lacks a hydrogen atom to contribute to a hydrogen bond, proline can only be stably accommodated within the first turn of an  $\alpha$ -helix. When present elsewhere, proline disrupts the conformation of the helix, producing a bend. Because of its small size, glycine also often induces bends in  $\alpha$ -helices.

**Figure 5-4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

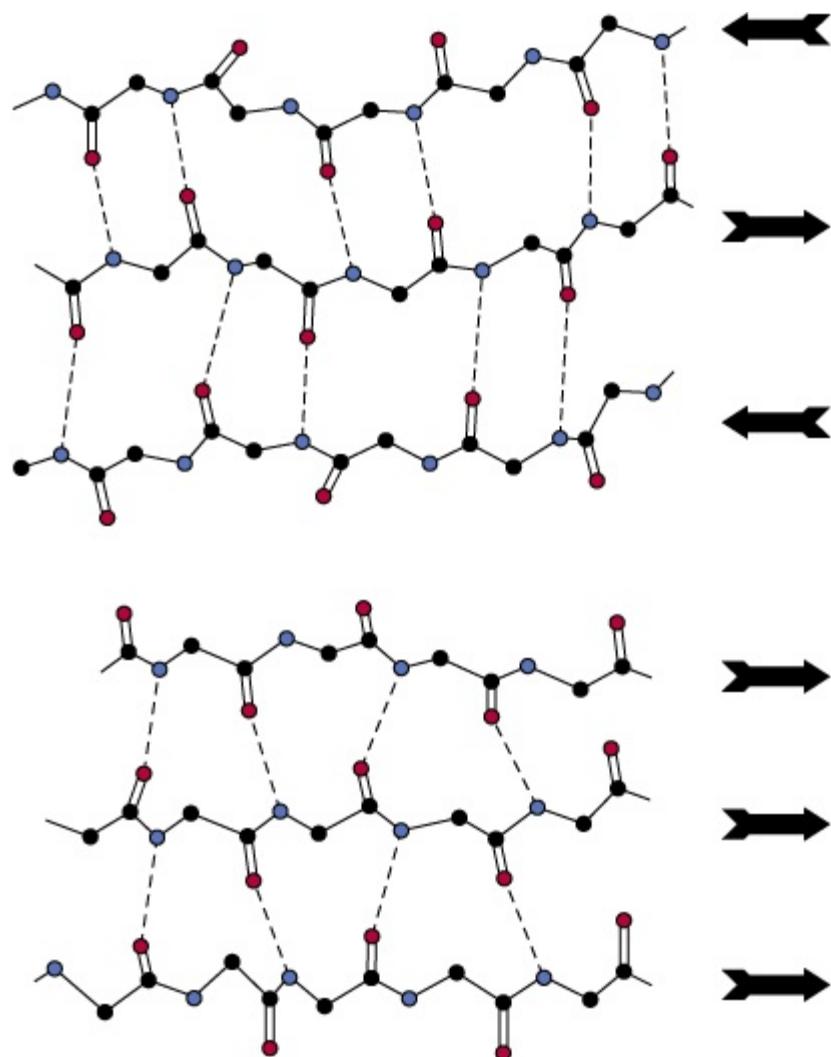
**Hydrogen bonds (dotted lines) formed between H and O atoms stabilize a polypeptide in an  $\alpha$ -helical conformation.** (Reprinted, with permission, from Haggis GH, et al., (1964), "Introduction to Molecular Biology". Science 146:1455-1456. Reprinted with permission from AAAS.)

Many  $\beta$  helices have predominantly hydrophobic R groups on one side of the axis of the helix and predominantly hydrophilic ones on the other. These **amphipathic helices** are well adapted to the formation of interfaces between polar and nonpolar regions such as the hydrophobic interior of a protein and its aqueous environment. Clusters of amphipathic helices can create a channel, or pore, that permits specific polar molecules to pass through hydrophobic cell membranes.

## Beta Sheet

The second (hence "beta") recognizable regular secondary structure in proteins is the  $\beta$  sheet. The amino acid residues of a  $\beta$  sheet, when viewed edge-on, form a zigzag or pleated pattern in which the R groups of adjacent residues point in opposite directions. Unlike the compact backbone of the  $\alpha$  helix, the peptide backbone of the  $\beta$  sheet is highly extended. But like the  $\alpha$  helix,  $\beta$  sheets derive much of their stability from hydrogen bonds between the carbonyl oxygens and amide hydrogens of peptide bonds. However, in contrast to the  $\alpha$  helix, these bonds are formed with adjacent segments of the  $\beta$  sheet (**Figure 5-5**).

**Figure 5-5**



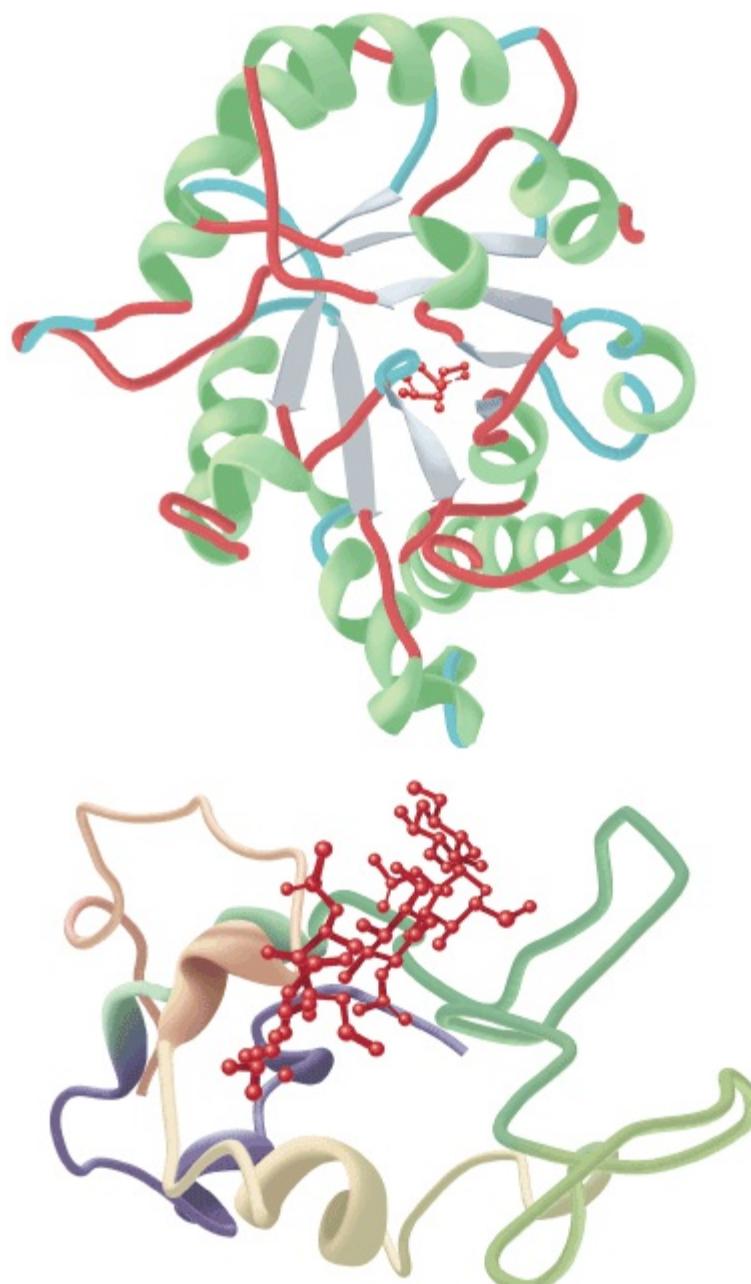
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Spacing and bond angles of the hydrogen bonds of antiparallel and parallel pleated sheets.** Arrows indicate the direction of each strand. Hydrogen bonds are indicated by dotted lines with the participating nitrogen atoms (hydrogen donors) and oxygen atoms (hydrogen acceptors) shown in blue and red, respectively. Backbone carbon atoms are shown in black. For clarity in presentation, R groups and hydrogen atoms are omitted. *Top:* Antiparallel sheet. Pairs of hydrogen bonds alternate between being close together and wide apart and are oriented approximately perpendicular to the polypeptide backbone. *Bottom:* Parallel sheet. The hydrogen bonds are evenly spaced but slant in alternate directions.

Interacting sheets can be arranged either to form a **parallel** sheet, in which the adjacent segments of the polypeptide chain proceed in the same direction amino to carboxyl, or an **antiparallel** sheet, in which they proceed in opposite directions (Figure 5–5). Either configuration permits the maximum number of hydrogen bonds between segments, or strands, of the sheet. Most sheets are not perfectly flat but tend to have a right-handed twist. Clusters of twisted strands of sheet form the core of many globular proteins (Figure 5–6). Schematic diagrams represent sheets as arrows that point in the amino to the carboxyl terminal direction.

**Figure 5–6**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

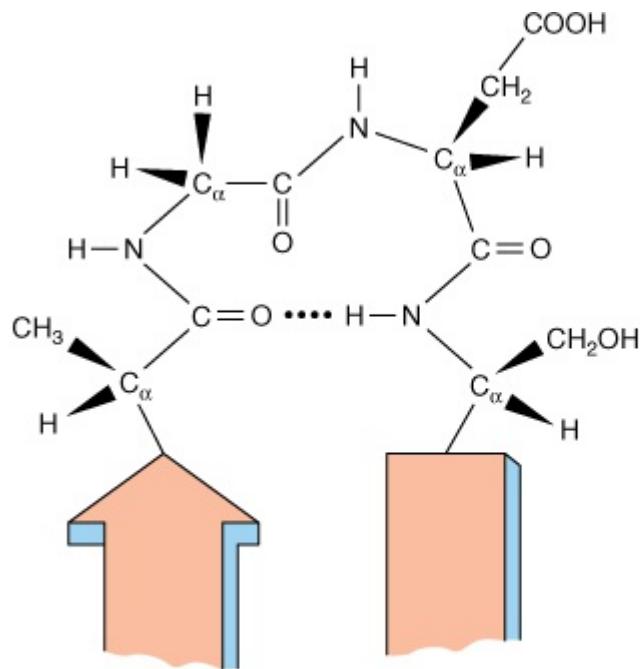
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Examples of the tertiary structure of proteins.** Top: The enzyme triose phosphate isomerase complexed with the substrate analog 2-phosphoglycerate (red). Note the elegant and symmetrical arrangement of alternating  $\beta$ -sheets (light blue) and  $\alpha$ -helices (green), with the  $\beta$ -sheets forming a  $\beta$ -barrel core surrounded by the helices. (Adapted from Protein Data Bank ID no. 1o5x.) Bottom: Lysozyme complexed with the substrate analog penta-N-acetyl chitopentaose (red). The color of the polypeptide chain is graded along the visible spectrum from purple (N-terminal) to tan (C-terminal). Notice how the concave shape of the domain forms a binding pocket for the pentasaccharide, the lack of  $\beta$ -sheet, and the high proportion of loops and bends. (Adapted from Protein Data Bank ID no. 1sfb.)

## Loops & Bends

Roughly half of the residues in a "typical" globular protein reside in  $\alpha$ -helices or  $\beta$ -sheets, and half in loops, turns, bends, and other extended conformational features. Turns and bends refer to short segments of amino acids that join two units of the secondary structure, such as two adjacent strands of an antiparallel  $\beta$ -sheet. A turn involves four aminoacyl residues, in which the first residue is hydrogen-bonded to the fourth, resulting in a tight 180° turn (Figure 5–7). Proline and glycine often are present in turns.

**Figure 5–7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**A turn that links two segments of anti-parallel sheet.** The dotted line indicates the hydrogen bond between the first and fourth amino acids of the four-residue segment Ala-Gly-Asp-Ser.

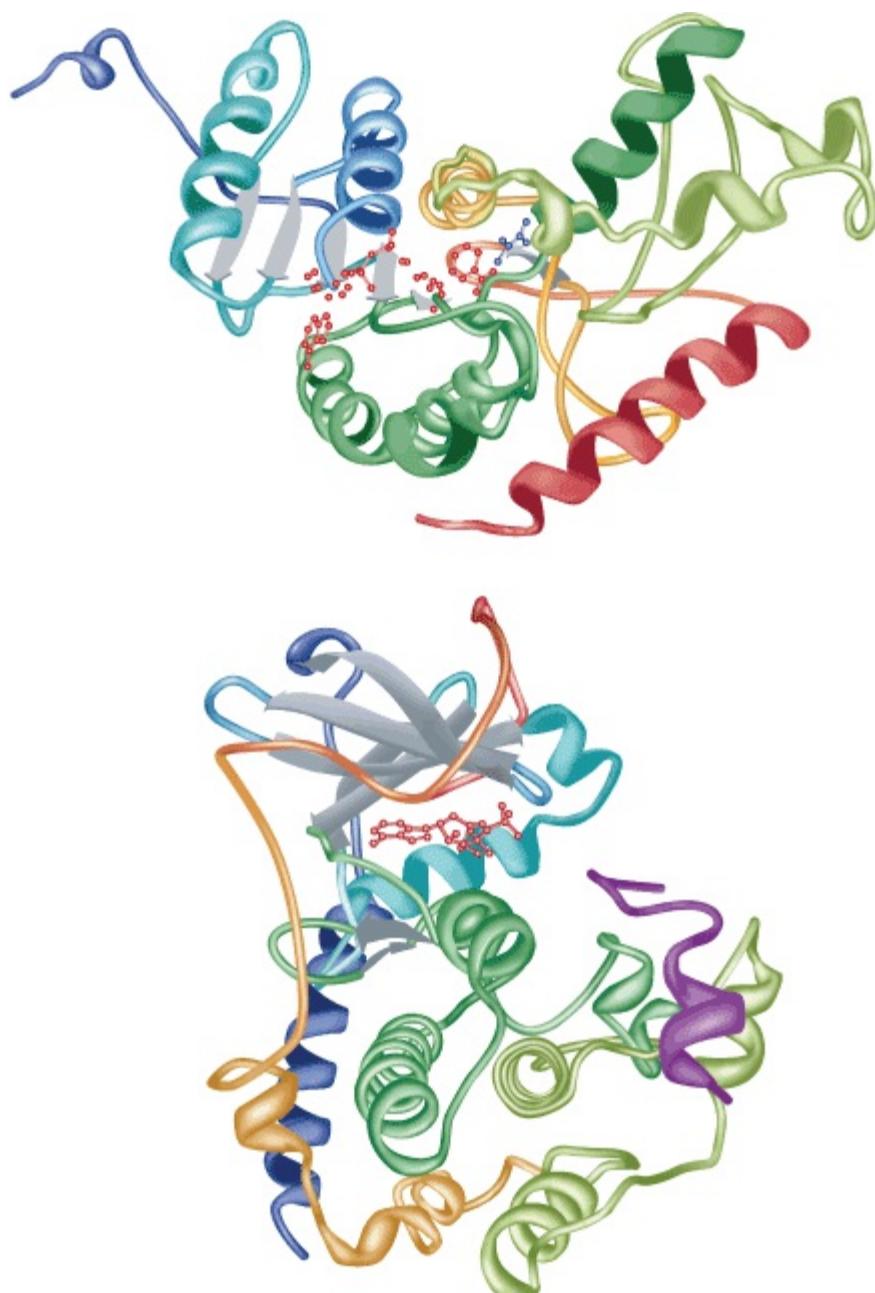
Loops are regions that contain residues beyond the minimum number necessary to connect adjacent regions of secondary structure. Irregular in conformation, loops nevertheless serve key biologic roles. For many enzymes, the loops that bridge domains responsible for binding substrates often contain aminoacyl residues that participate in catalysis. **Helix-loop-helix motifs** provide the oligonucleotide-binding portion of many DNA-binding proteins such as repressors and transcription factors. Structural motifs such as the helix-loop-helix motif that are intermediate in scale between secondary and tertiary structures are often termed **supersecondary structures**. Since many loops and bends reside on the surface of proteins and are thus exposed to solvent, they constitute readily accessible sites, or **epitopes**, for recognition and binding of antibodies.

While loops lack apparent structural regularity, many adopt a specific conformation stabilized through hydrogen bonding, salt bridges, and hydrophobic interactions with other portions of the protein. However, not all portions of proteins are necessarily ordered. Proteins may contain "disordered" regions, often at the extreme amino or carboxyl terminal, characterized by high conformational flexibility. In many instances, these disordered regions assume an ordered conformation upon binding of a ligand. This structural flexibility enables such regions to act as ligand-controlled switches that affect protein structure and function.

## Tertiary & Quaternary Structure

The term "tertiary structure" refers to the entire three-dimensional conformation of a polypeptide. It indicates, in three-dimensional space, how secondary structural features—helices, sheets, bends, turns, and loops—assemble to form domains and how these domains relate spatially to one another. A **domain** is a section of the protein structure sufficient to perform a particular chemical or physical task such as binding of a substrate or other ligand. Most domains are modular in nature, and contiguous in both primary sequence and three-dimensional space (**Figure 5–8**). Simple proteins, particularly those that interact with a single substrate, such as lysozyme or triose phosphate isomerase (**Figure 5–6**) and the oxygen storage protein myoglobin (Chapter 6), often consist of a single domain. By contrast, lactate dehydrogenase is comprised of two domains, an N-terminal NAD<sup>+</sup>-binding domain and a C-terminal binding domain for the second substrate, pyruvate (**Figure 5–8**). Lactate dehydrogenase is one of the family of oxidoreductases that share a common N-terminal NAD(P)<sup>+</sup>-binding domain known as the **Rossmann fold**. By fusing the Rossmann fold domain to a variety of C-terminal domains, a large family of oxidoreductases have evolved that utilize NAD(P)<sup>+</sup>/NAD(P)H for the oxidation and reduction of a wide range of metabolites. Examples include alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, malate dehydrogenase, quinone oxidoreductase, 6-phosphogluconate dehydrogenase, D-glycerate dehydrogenase, formate dehydrogenase, and 3-, 20 -hydroxysteroid dehydrogenase.

**Figure 5–8**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

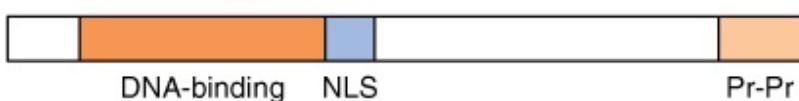
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Polypeptides containing two domains.** Top: Shown is the three-dimensional structure of a monomer unit of the tetrameric enzyme lactate dehydrogenase with the substrates NADH (red) and pyruvate (blue) bound. Not all bonds in NADH are shown. The color of the polypeptide chain is graded along the visible spectrum from blue (N-terminal) to orange (C-terminal). Note how the N-terminal portion of the polypeptide forms a contiguous domain, encompassing the left portion of the enzyme, responsible for binding NADH. Similarly, the C-terminal portion forms a contiguous domain responsible for binding pyruvate. (Adapted from Protein Data Bank ID no. 3ldh.) Bottom: Shown is the three-dimensional structure of the catalytic subunit of the cAMP-dependent protein kinase (Chapter 42) with the substrate analogs ADP (red) and peptide (purple) bound. The color of the polypeptide chain is graded along the visible spectrum from blue (N-terminal) to orange (C-terminal). Protein kinases transfer the  $\gamma$ -phosphate group of ATP to protein and peptide substrates (Chapter 9). Note how the N-terminal portion of the polypeptide forms a contiguous domain rich in  $\beta$ -sheet that binds ADP. Similarly, the C-terminal portion forms a contiguous,  $\alpha$ -helix-rich domain responsible for binding the peptide substrate. (Adapted from Protein Data Bank ID no. 1jbp.)

Not all domains bind substrates. Hydrophobic membrane domains anchor proteins to membranes or enable them to span membranes. Localization sequences target proteins to specific subcellular or extracellular locations such as the nucleus, mitochondria, secretory vesicles, etc. Regulatory domains trigger changes in protein function in response to the binding of allosteric effectors or covalent modifications (Chapter 9). Combining domain modules provides a facile route for generating proteins of great structural complexity and functional sophistication (Figure 5–9).

**Figure 5–9**

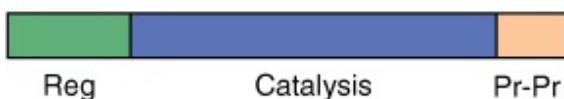
## Forkhead transcription factor



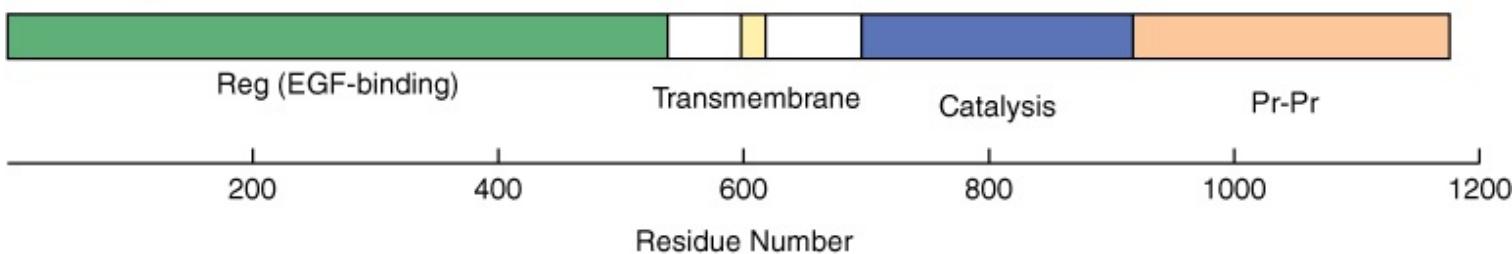
## 6-Phosphofructo-2-kinase/Fructose-2,6-bisphosphatase



## Phenylalanine Hydroxylase



## EGF Receptor



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Some multidomain proteins.** The rectangles represent the polypeptide sequences of a forkhead transcription factor; 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, a bifunctional enzyme whose activities are controlled in a reciprocal fashion by allosteric effectors and covalent modification (Chapter 20); phenylalanine hydroxylase (Chapters 27 and 29), whose activity is stimulated by phosphorylation of its regulatory domain; and the epidermal growth factor (EGF) receptor (Chapter 41), a transmembrane protein whose intracellular protein kinase domain is regulated via the binding of the peptide hormone EGF to its extracellular domain. Regulatory domains are colored green, catalytic domains dark blue and light blue, protein-protein interaction domains light orange, DNA binding domains dark orange, nuclear localization sequences medium blue, and transmembrane domains yellow. The kinase and bisphosphatase activities of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase are catalyzed by the N- and C-terminal proximate catalytic domains, respectively.

Proteins containing multiple domains can also be assembled through the association of multiple polypeptides, or protomers. Quaternary structure defines the polypeptide composition of a protein and, for an oligomeric protein, the spatial relationships between its protomers or subunits. **Monomeric** proteins consist of a single polypeptide chain. **Dimeric** proteins contain two polypeptide chains. **Homodimers** contain two copies of the same polypeptide chain, while in a **heterodimer** the polypeptides differ. Greek letters ( $\alpha$ ,  $\beta$ ,  $\gamma$ , etc) are used to distinguish different subunits of a hetero-oligomeric protein, and subscripts indicate the number of each subunit type. For example,  $\alpha_4$  designates a homotetrameric protein, and  $\alpha_2\beta_2$  a protein with five subunits of three different types.

Since even small proteins contain many thousands of atoms, depictions of protein structure that indicate the position of every atom are generally too complex to be readily interpreted. Simplified schematic diagrams thus are used to depict the key features of a protein's tertiary and quaternary structure. Ribbon diagrams (Figures 5–6 and 5–8) trace the conformation of the polypeptide backbone, with cylinders and arrows indicating regions of  $\alpha$ -helix and  $\beta$ -sheet, respectively. In an even simpler representation, line segments that link the  $\alpha$ -carbons indicate the path of the polypeptide backbone. These schematic diagrams often include the side chains of selected amino acids that emphasize specific structure-function relationships.

## MULTIPLE FACTORS STABILIZE THE TERTIARY & QUATERNARY STRUCTURE

Higher orders of protein structure are stabilized primarily—and often exclusively—by noncovalent interactions. Principal among these are hydrophobic interactions that drive most hydrophobic amino acid side chains into the interior of the protein, shielding them from water. Other significant contributors include hydrogen bonds and salt bridges between the carboxylates of aspartic and glutamic acid and the oppositely charged side chains of protonated lysyl, arginyl, and histidyl residues. While individually weak relative to a typical covalent bond of 80–120 kcal/mol, collectively these numerous interactions confer a high degree of stability to the biologically functional conformation of a protein, just as a Velcro fastener harnesses the cumulative strength of a multitude of tiny plastic loops and hooks.

Some proteins contain covalent disulfide (S—S) bonds that link the sulfhydryl groups of cysteinyl residues. Formation of disulfide bonds involves oxidation of the cysteinyl sulfhydryl groups and requires oxygen. Intrapolyptide disulfide bonds further enhance the stability of the folded conformation of a peptide, while interpolyptide disulfide bonds stabilize the quaternary structure of certain oligomeric proteins.

## THREE-DIMENSIONAL STRUCTURE IS DETERMINED BY X-RAY CRYSTALLOGRAPHY OR BY NMR SPECTROSCOPY

### X-Ray Crystallography

Following the solution in 1960 by John Kendrew of the three-dimensional structure of myoglobin, x-ray crystallography revealed the structures of thousands of biological macromolecules ranging from proteins to many oligonucleotides and a few viruses. For the solution of its structure by x-ray crystallography, a protein is first precipitated under conditions that form large, well-ordered crystals. To establish appropriate conditions, crystallization trials use a few microliters of protein solution and a matrix of variables (temperature, pH, presence of salts or organic solutes such as polyethylene glycol) to establish optimal conditions for crystal formation. Crystals mounted in quartz capillaries are first irradiated with monochromatic x-rays of approximate wavelength 0.15 nm to confirm that they are protein, not salt. Protein crystals may then be frozen in liquid nitrogen for subsequent collection of a high-resolution data set. The patterns formed by the x-rays that are diffracted by the atoms in their path are recorded on a photographic plate or its computer equivalent as a circular pattern of spots of varying intensity. The data inherent in these spots are then analyzed using a mathematical approach termed a *Fourier synthesis*, which summates wave functions. The wave amplitudes are related to spot intensity, but since the waves are not in phase, the relationship between their phases must next be determined.

The traditional approach to solution of the "phase problem" employs **isomorphous displacement**. Prior to irradiation, an atom with a distinctive x-ray "signature" is introduced into a crystal at known positions in the primary structure of the protein. Heavy atom isomorphous displacement generally uses mercury or uranium, which bind to cysteine residues. An alternative approach uses the expression of plasmid-encoded recombinant proteins in which selenium replaces the sulfur of methionine. Expression uses a bacterial host auxotrophic for methionine biosynthesis and a defined medium in which selenomethionine replaces methionine. Alternatively, if the unknown structure is similar to one that has already been solved, **molecular replacement** on an existing model provides an attractive way to phase the data without the use of heavy atoms. Finally, the results from the phasing and Fourier summations provide an electron density profile or three-dimensional map of how the atoms are connected or related to one another.

### Laue X-Ray Crystallography

The ability of some crystallized enzymes to catalyze chemical reactions strongly suggests that structures determined by crystallography are indeed representative of the structures present in the free solution. Classic crystallography provides, however, an essentially static picture of a protein that may undergo significant structural changes *in vivo*, such as those that accompany enzymic catalysis. The Laue approach uses diffraction of *polychromatic* x-rays, and many crystals. The time-consuming process of rotating the crystal in the x-ray beam is avoided, which permits the use of extremely short exposure times. Detection of the motions of residues or domains of an enzyme during catalysis uses crystals that contain an inactive or "caged" substrate analog. An intense flash of visible light cleaves the caged precursor to release free substrate and initiate catalysis in a precisely controlled manner. Using this approach, data can be collected over time periods as short as a few nanoseconds.

### Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy, a powerful complement to x-ray crystallography, measures the absorbance of radio frequency electromagnetic energy by certain atomic nuclei. "NMR-active" isotopes of biologically relevant elements include  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$ . The frequency, or chemical shift, at which a particular nucleus absorbs energy is a function of both the functional group within which it resides and the proximity of other NMR-active nuclei. Once limited to metabolites and relatively small macromolecules,  $< 30 \text{ kDa}$ , today proteins and protein complexes of  $> 100 \text{ kDa}$  can be analyzed by NMR. Two-dimensional NMR spectroscopy permits a three-dimensional representation of a protein to be constructed by determining the proximity of these nuclei to one another. NMR spectroscopy analyzes proteins in aqueous solution. Not only does this obviate the need to form crystals (a particular advantage when dealing with difficult to crystallize membrane proteins), it renders possible real-time observation of the changes in conformation that accompany ligand binding or catalysis. It also offers the possibility of perhaps one day being able to observe the structure and dynamics of proteins (and metabolites) within living cells.

### Molecular Modeling

A valuable adjunct to the empirical determination of the three-dimensional structure of proteins is the use of computer technology for molecular modeling. When the three-dimensional structure is known, **molecular dynamics** programs can be used to simulate the conformational dynamics of a protein and the manner in which factors such as temperature, pH, ionic strength, or amino acid substitutions influence these motions. **Molecular docking** programs simulate the interactions that take place when a protein encounters a substrate, inhibitor, or other ligand. Virtual screening for molecules likely to interact with key sites on a protein of biomedical interest is extensively used to facilitate the discovery of new drugs.

Molecular modeling is also employed to infer the structure of proteins for which x-ray crystallographic or NMR structures are not yet available. Secondary structure algorithms weigh the propensity of specific residues to become incorporated into  $\alpha$ -helices or  $\beta$ -sheets in previously studied proteins to predict the secondary structure of other polypeptides. In **homology modeling**, the known three-dimensional structure of a protein is used as a template upon which to erect a model of the probable structure of a related protein. Scientists are working to devise computer programs that will reliably predict the three-dimensional conformation of a protein directly from its primary sequence, thereby permitting determination of the structures of the many unknown proteins for which templates currently are lacking.

## PROTEIN FOLDING

Proteins are conformationally dynamic molecules that can fold into their functionally competent conformation in a time frame of milliseconds, and often can refold if their conformation becomes disrupted, or denatured. How is this remarkable process achieved? Folding into the native state does not involve a haphazard search of all possible structures. Denatured proteins are not just random coils. Native contacts are favored, and regions of the native structure persist even in the denatured state. Discussed below are factors that facilitate folding and refolding, and current concepts and proposed mechanisms based on more than 40 years of largely *in vitro* experimentation.

### Native Conformation of a Protein Is Thermodynamically Favored

The number of distinct combinations of phi and psi angles specifying potential conformations of even a relatively small—15 kDa—polypeptide is unbelievably vast. Proteins are guided through this vast labyrinth of possibilities by thermodynamics. Since the biologically relevant—or native—conformation of a protein generally is the one that is most energetically favored, knowledge of the native conformation is specified in the primary sequence. However, if one were to wait for a polypeptide to find its native conformation by random exploration of all possible conformations, the process would require billions of years to complete. Clearly, in nature, protein folding takes place in a more orderly and guided fashion.

### Folding Is Modular

Protein folding generally occurs via a stepwise process. In the first stage, as the newly synthesized polypeptide emerges from the ribosome, short segments fold into secondary structural units that provide local regions of organized structure. Folding is now reduced to the selection of an appropriate arrangement of this relatively small number of secondary structural elements. In the second stage, the hydrophobic regions segregate into the interior of the protein away from solvent, forming a "molten globule," a partially folded polypeptide in which the modules of the secondary structure rearrange until the mature conformation of the protein is attained. This process is orderly, but not rigid. Considerable flexibility exists in the ways and in the order in which elements of secondary structure can be rearranged. In general, each element of the secondary or super-secondary structure facilitates proper folding by directing the folding process toward the native conformation and away from unproductive alternatives. For oligomeric proteins, individual protomers tend to fold before they associate with other subunits.

### Auxiliary Proteins Assist Folding

Under appropriate laboratory conditions, many proteins will spontaneously refold after being **denatured** (ie, unfolded) by treatment with acid or base, chaotropic agents, or detergents. However, refolding under these conditions is slow—minutes to hours. Moreover, many proteins fail to spontaneously refold *in vitro*. Instead they form insoluble **aggregates**, disordered complexes of unfolded or partially folded polypeptides held together predominantly by hydrophobic interactions. Aggregates represent unproductive dead ends in the folding process. Cells employ auxiliary proteins to speed the process of folding and to guide it toward a productive conclusion.

### Chaperones

**Chaperone** proteins participate in the folding of over half of all mammalian proteins. The hsp70 (70 kDa heat shock protein) family of chaperones binds short sequences of hydrophobic amino acids that emerge while a new polypeptide is being synthesized, shielding them from solvent. Chaperones prevent aggregation, thus providing an opportunity for the formation of appropriate secondary structural elements and their subsequent coalescence into a molten globule. The hsp60 family of chaperones, sometimes called **chaperonins**, differ in sequence and structure from hsp70 and its homologs. Hsp60 acts later in the folding process, often together with an hsp70 chaperone. The central cavity of the donut-shaped hsp60 chaperone provides a sheltered environment in which a polypeptide can fold until all hydrophobic regions are buried in its interior, thus preempting any tendency toward aggregation.

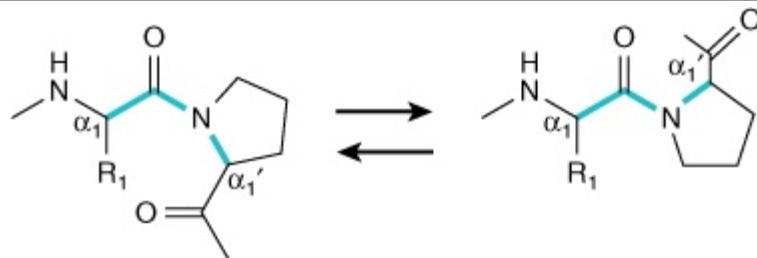
### Protein Disulfide Isomerase

Disulfide bonds between and within polypeptides stabilize tertiary and quaternary structures. However, disulfide bond formation is nonspecific. Under oxidizing conditions, a given cysteine can form a disulfide bond with the —SH of any accessible cysteinyl residue. By catalyzing disulfide exchange, the rupture of an S—S bond and its reformation with a different partner cysteine, protein disulfide isomerase facilitates the formation of disulfide bonds that stabilize a protein's native conformation.

### Proline-*cis*, *trans*-Isomerase

All X-Pro peptide bonds—where X represents any residue—are synthesized in the *trans* configuration. However, of the X-Pro bonds of mature proteins, approximately 6% are *cis*. The *cis* configuration is particularly common in  $\alpha$ -turns. Isomerization from *trans* to *cis* is catalyzed by the enzyme proline-*cis*, *trans*-isomerase (**Figure 5-10**).

**Figure 5-10**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Isomerization of the N-<sub>1</sub> prolyl peptide bond from a *cis* to a *trans* configuration relative to the backbone of the polypeptide.**

### Folding Is a Dynamic Process

Proteins are conformationally dynamic molecules that can fold and unfold hundreds or thousands of times in their lifetime. How do proteins, once unfolded, refold and restore their functional conformation? First, unfolding rarely leads to the complete randomization of the polypeptide chain inside the cell. Unfolded proteins generally retain a number of contacts and regions of the secondary structure that facilitate the refolding process. Second, chaperone proteins can "rescue" unfolded proteins that have become thermodynamically trapped in a misfolded dead end by unfolding hydrophobic regions and providing a second chance to fold productively. Glutathione can reduce inappropriate disulfide bonds that may be formed upon exposure to oxidizing agents such as O<sub>2</sub>, hydrogen peroxide, or superoxide (Chapter 52).

## PERTURBATION OF PROTEIN CONFORMATION MAY HAVE PATHOLOGIC CONSEQUENCES

### Prions

The transmissible spongiform encephalopathies, or **prion diseases**, are fatal neurodegenerative diseases characterized by spongiform changes, astrocytic gliomas, and neuronal loss resulting from the deposition of insoluble protein aggregates in neural cells. They include Creutzfeldt-Jakob disease in humans, scrapie in sheep, and bovine spongiform encephalopathy (mad cow disease) in cattle. A variant form of Creutzfeldt-Jacob disease (vCJD) that afflicts younger patients is associated with early-onset psychiatric and behavioral disorders. Prion diseases may manifest themselves as infectious, genetic, or sporadic disorders. Because no viral or bacterial gene encoding the pathologic prion protein could be identified, the source and mechanism of transmission of prion disease long remained elusive.

Today it is recognized that prion diseases are protein conformation diseases transmitted by altering the conformation, and hence the physical properties, of proteins endogenous to the host. Human prion-related protein (PrP), a glycoprotein encoded on the short arm of chromosome 20, normally is monomeric and rich in  $\alpha$ -helix. Pathologic prion proteins serve as the templates for the conformational transformation of normal PrP, known as PrP<sub>c</sub>, into PrP<sub>s</sub>. PrP<sub>s</sub> is rich in  $\beta$ -sheet with many hydrophobic aminoacyl side chains exposed to solvent. As each new PrP<sub>s</sub> molecule is formed, it triggers the production of yet more pathologic variants in a conformational chain reaction. Because PrP<sub>s</sub> molecules associate strongly with one other through their exposed hydrophobic regions, the accumulating PrP<sub>s</sub> units coalesce to form insoluble protease-resistant aggregates. Since one pathologic prion or prion-related protein can serve as template for the conformational transformation of many times its number of PrP<sub>c</sub> molecules, prion diseases can be transmitted by the protein alone without involvement of DNA or RNA.

### Alzheimer's Disease

Refolding or misfolding of another protein endogenous to human brain tissue,  $\beta$ -amyloid, is a prominent feature of Alzheimer's disease. While the main cause of Alzheimer's disease remains elusive, the characteristic senile plaques and neurofibrillary bundles contain aggregates of the protein  $\beta$ -amyloid, a 4.3 kDa polypeptide produced by proteolytic cleavage of a larger protein known as amyloid precursor protein. In Alzheimer's disease patients, levels of  $\beta$ -amyloid become elevated, and this protein undergoes a conformational transformation from a soluble  $\alpha$ -helix-rich state to a state rich in  $\beta$ -sheet and prone to self-aggregation. Apolipoprotein E has been implicated as a potential mediator of this conformational transformation.

### Beta-Thalassemias

Thalassemias are caused by genetic defects that impair the synthesis of one of the polypeptide subunits of hemoglobin (Chapter 6). During the burst of hemoglobin synthesis that occurs during erythrocyte development, a specific chaperone called  $\alpha$ -hemoglobin-stabilizing protein (AHSP) binds to free hemoglobin  $\alpha$ -subunits awaiting incorporation into the hemoglobin multimer. In the absence of this chaperone, free  $\alpha$ -hemoglobin subunits aggregate, and the resulting precipitate has cytotoxic effects on the developing erythrocyte.

Investigations using genetically modified mice suggest a role for AHSP in modulating the severity of  $\beta$ -thalassemia in human subjects.

## COLLAGEN ILLUSTRATES THE ROLE OF POSTTRANSLATIONAL PROCESSING IN PROTEIN MATURATION

### Protein Maturation Often Involves Making & Breaking of Covalent Bonds

The maturation of proteins into their final structural state often involves the cleavage or formation (or both) of covalent bonds, a process of **post-translational modification**. Many polypeptides are initially synthesized as larger precursors called **proteins**. The "extra" polypeptide segments in these proproteins often serve as leader sequences that target a polypeptide to a particular organelle or facilitate its passage through a membrane. Other segments ensure that the potentially harmful activity of a protein such as the protease trypsin and chymotrypsin remains inhibited until these proteins reach their final destination. However, once these transient requirements are fulfilled and the now superfluous peptide regions are removed by selective proteolysis. Other covalent modifications may take place that add new chemical functionalities to a protein. The maturation of collagen illustrates both of these processes.

### Collagen Is a Fibrous Protein

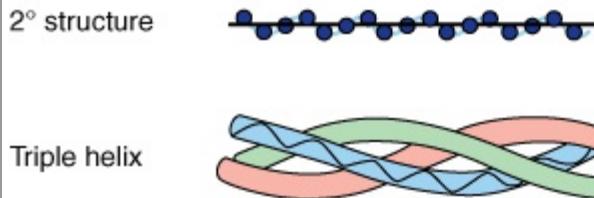
Collagen is the most abundant of the fibrous proteins that constitute more than 25% of the protein mass in the human body. Other prominent fibrous proteins include keratin and myosin. These fibrous proteins represent a primary source of structural strength for cells (ie, the cytoskeleton) and tissues. Skin derives its strength and flexibility from an intertwined mesh of collagen and keratin fibers, while bones and teeth are buttressed by an underlying network of collagen fibers analogous to steel strands in reinforced concrete. Collagen also is present in connective tissues such as ligaments and tendons. The high degree of tensile strength required to fulfill these structural roles requires elongated proteins characterized by repetitive amino acid sequences and a regular secondary structure.

### Collagen Forms a Unique Triple Helix

Tropocollagen, the repeating unit of a mature collagen fiber, consists of three collagen polypeptides, each containing about 1000 amino acids, bundled together in a unique conformation, the collagen triple helix (**Figure 5-11**). A mature collagen fiber forms an elongated rod with an axial ratio of about 200. Three intertwined polypeptide strands, which twist to the left, wrap around one another in a right-handed fashion to form the collagen triple helix. The opposing handedness of this superhelix and its component polypeptides makes the collagen triple helix highly resistant to unwinding—a principle also applied to the steel cables of suspension bridges. A collagen triple helix has 3.3 residues per turn and a rise per residue nearly twice that of an  $\alpha$ -helix. The R groups of each polypeptide strand of the triple helix pack so closely that, in order to fit, one of the three must be H. Thus, every third amino acid residue in collagen is a glycine residue. Staggering of the three strands provides appropriate positioning of the requisite glycines throughout the helix. Collagen is also rich in proline and hydroxyproline, yielding a repetitive Gly-X-Y pattern (**Figure 5-11**) in which Y generally is proline or hydroxyproline.

**Figure 5-11**

Amino acid sequence    -Gly – X – Y – Gly – X – Y – Gly – X – Y –



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary, secondary, and tertiary structures of collagen.

Collagen triple helices are stabilized by hydrogen bonds between residues in *different* polypeptide chains, a process helped by the hydroxyl groups of hydroxyprolyl residues. Additional stability is provided by covalent cross links formed between modified lysyl residues both within and between polypeptide chains.

### Collagen Is Synthesized as a Larger Precursor

Collagen is initially synthesized as a larger precursor polypeptide, procollagen. Numerous prolyl and lysyl residues of procollagen are hydroxylated by prolyl hydroxylase and lysyl hydroxylase, enzymes that require ascorbic acid (vitamin C; see Chapters 27 and 44). Hydroxyprolyl and hydroxylysyl residues provide additional hydrogen bonding capability that stabilizes the mature protein. In addition, glucosyl and galactosyl transferases attach glucosyl or galactosyl residues to the hydroxyl groups of specific hydroxylysyl residues.

The central portion of the precursor polypeptide then associates with other molecules to form the characteristic triple helix. This process is accompanied by the removal of the globular amino terminal and carboxyl terminal extensions of the precursor polypeptide by selective proteolysis. Certain lysyl residues are modified by lysyl oxidase, a copper-containing protein that converts  $\alpha$ -amino groups to aldehydes. The aldehydes can either undergo an aldol condensation to form a C=C double bond or to form a Schiff base (eneimine) with the  $\alpha$ -amino group of an unmodified lysyl residue, which is subsequently reduced to form a C=N single bond. These covalent bonds cross-link the individual polypeptides and imbue the fiber with exceptional strength and rigidity.

### Nutritional & Genetic Disorders Can Impair Collagen Maturation

The complex series of events in collagen maturation provide a model that illustrates the biologic consequences of incomplete polypeptide maturation. The best-known defect in collagen biosynthesis is scurvy, a result of a dietary deficiency of vitamin C required by prolyl and lysyl hydroxylases. The resulting deficit in the number of hydroxyproline and hydroxylysine residues undermines the conformational stability of collagen fibers, leading to bleeding gums, swelling joints, poor wound healing, and ultimately death. Menkes' syndrome, characterized by kinky hair and growth retardation, reflects a dietary deficiency of the copper required by lysyl oxidase, which catalyzes a key step in the formation of the covalent cross-links that strengthen collagen fibers.

Genetic disorders of collagen biosynthesis include several forms of osteogenesis imperfecta, characterized by fragile bones. In the Ehlers-Danlos syndrome, a group of connective tissue disorders that involve impaired integrity of supporting structures, defects in the genes that encode collagen-1, procollagen N-peptidase, or lysyl hydroxylase result in mobile joints and skin abnormalities (see also Chapter 48).

## SUMMARY

- Proteins may be classified based on their solubility, shape, or function or on the presence of a prosthetic group, such as heme.
- The gene-encoded primary structure of a polypeptide is the sequence of its amino acids. Its secondary structure results from folding of polypeptides into hydrogen-bonded motifs such as the  $\alpha$  helix, the  $\beta$  pleated sheet,  $\beta$  bends, and loops. Combinations of these motifs can form supersecondary motifs.
- Tertiary structure concerns the relationships between secondary structural domains. Quaternary structure of proteins with two or more polypeptides (oligomeric proteins) concerns the spatial relationships between various types of polypeptides.
- Primary structures are stabilized by covalent peptide bonds. Higher orders of structure are stabilized by weak forces—multiple hydrogen bonds, salt (electrostatic) bonds, and association of hydrophobic R groups.
- The phi ( $\phi$ ) angle of a polypeptide is the angle about the C $\alpha$ —N bond; the psi ( $\psi$ ) angle is that about the C $\alpha$ —C $\beta$  bond. Most combinations of phi-psi angles are disallowed due to steric hindrance. The phi-psi angles that form the  $\alpha$  helix and the  $\beta$  sheet fall within the lower and upper left-hand quadrants of a Ramachandran plot, respectively.
- Protein folding is a poorly understood process. Broadly speaking, short segments of newly synthesized polypeptide fold into secondary structural units. Forces that bury hydrophobic regions from solvent then drive the partially folded polypeptide into a "molten globule" in which the modules of the secondary structure are rearranged to give the native conformation of the protein.
- Proteins that assist folding include protein disulfide isomerase, proline-*cis*, *trans*-isomerase, and the chaperones that participate in the folding of over half of mammalian proteins. Chaperones shield newly synthesized polypeptides from solvent and provide an environment for elements of secondary structure to emerge and coalesce into molten globules.
- X-Ray crystallography and NMR are key techniques used to study higher orders of protein structure.
- Prions—protein particles that lack nucleic acid—cause fatal transmissible spongiform encephalopathies such as Creutzfeldt–Jakob disease, scrapie, and bovine spongiform encephalopathy. Prion diseases involve an altered secondary-tertiary structure of a naturally occurring protein, PrP $\text{c}$ . When PrP $\text{c}$  interacts with its pathologic isoform PrP $\text{Sc}$ , its conformation is transformed from a predominantly  $\alpha$ -helical structure to the  $\beta$ -sheet structure characteristic of PrP $\text{Sc}$ .
- Collagen illustrates the close linkage between protein structure and biologic function. Diseases of collagen maturation include Ehlers–Danlos syndrome and the vitamin C deficiency disease scurvy.

## REFERENCES

- Caughey B, Baron GS, Chesebro B, et al: Getting a grip on prions: oligomers, amyloids, and pathological membrane interactions. *Annu Rev Biochem* 2009;78:177.
- Chiti F, Dobson CM: Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 2006;75:519.
- Foster MP, McElroy CA, Amero CD: Solution NMR of large molecules and assemblies. *Biochemistry* 2007;46:331.
- Gothel SF, Marahiel MA: Peptidyl-prolyl *cis-trans* isomerasers, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci* 1999;55:423.
- Hardy J: Toward Alzheimer therapies based on genetic knowledge. *Annu Rev Med* 2004;55:15.
- Hartl FU, Hayer-Hartl M: Converging concepts of protein folding in vitro and in vivo. *Nat Struct Biol* 2009;16:574.
- Ho BK, Thomas A, Brasseur R: Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the  $\alpha$ -helix. *Protein Sci* 2003;12:2508.
- Hristova K, Wimley WC, Mishra VK, et al: An amphipathic alpha-helix at a membrane interface: a structural study using a novel X-ray diffraction method. *J Mol Biol* 1999;290:99.
- Irani DN, Johnson RT: Diagnosis and prevention of bovine spongiform encephalopathy and variant Creutzfeldt-Jakob disease. *Annu Rev Med* 2003;54:305.
- Jorgensen WL: The many roles of computation in drug discovery. *Science* 2004;303:1813.
- Khare SD, Dokholyan NV: Molecular mechanisms of polypeptide aggregation in human disease. *Curr Protein Pept Sci* 2007;8:573.
- Kim J, Holtzman DM: Prion-like behavior of amyloid- $\beta$ . *Science* 2010;330:918.
- Kong Y, Zhou S, Kihm AJ, et al: Loss of alpha-hemoglobin-stabilizing protein impairs erythropoiesis and exacerbates beta-thalassemia. *J Clin Invest* 2004;114:1457.
- Myllyharju J: Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol* 2003;22:15.
- Rider MH, Bertrand L, Vertommen D, et al: 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: head-to-head with a bifunctional enzyme that controls glycolysis. *Biochem J* 2004;381:561.
- Shoulders MD, Raines RT: Collagen structure and stability. *Annu Rev Biochem* 2009; 78:929.
- Stoddard BL, Cohen BE, Brubaker M, et al: Millisecond Laue structures of an enzyme-product complex using photocaged substrate analogs. *Nat Struct Biol* 1998;5:891.
- Wegezyn RD, Deuerling E: Molecular guardians for newborn proteins: ribosome-associated chaperones and their role in protein folding. *Cellular Mol Life Sci* 2005;62:2727.
- Young JC, Moarefi I, Hartl FU: *Hsp90*: a specialized but essential protein-folding tool. *J Cell Biol* 2001;154:267.

## OBJECTIVES

After studying this chapter, you should be able to:

- Describe the most important structural similarities and differences between myoglobin and hemoglobin.
- Sketch binding curves for the oxygenation of myoglobin and hemoglobin.
- Identify the covalent linkages and other close associations between heme and globin in oxymyoglobin and oxyhemoglobin.
- Explain why the physiologic function of hemoglobin requires that its O<sub>2</sub>-binding curve be sigmoidal rather than hyperbolic.
- Explain the role of a hindered environment on the ability of hemoglobin to bind carbon monoxide.
- Define P<sub>50</sub> and indicate its significance in oxygen transport and delivery.
- Describe the structural and conformational changes in hemoglobin that accompany its oxygenation and subsequent deoxygenation.
- Explain the role of 2,3-bisphosphoglycerate (BPG) in oxygen binding and delivery.
- Outline the role of hemoglobin in CO<sub>2</sub> and proton transport, and describe accompanying changes in the pK<sub>a</sub> of the relevant imidazolium group.
- Describe the structural consequences to HbS of lowering pO<sub>2</sub>.
- Identify the metabolic defect that occurs as a consequence of - and - thalassemias.

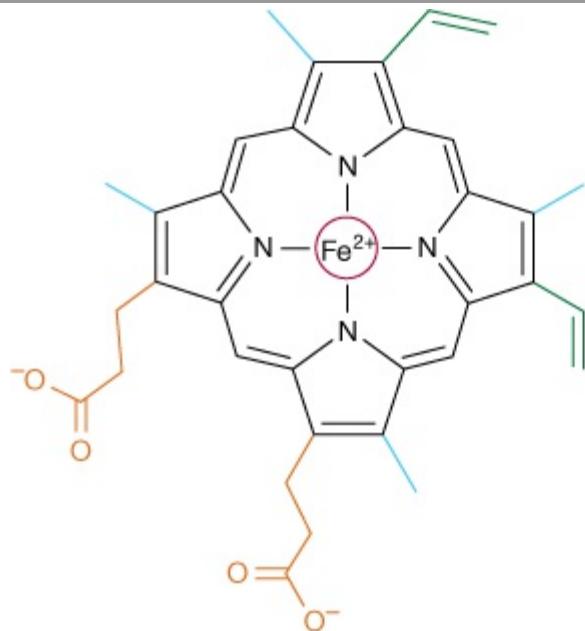
## BIOMEDICAL IMPORTANCE

The heme proteins myoglobin and hemoglobin maintain a supply of oxygen essential for oxidative metabolism. Myoglobin, a monomeric protein of red muscle, stores oxygen as a reserve against oxygen deprivation. Hemoglobin, a tetrameric protein of erythrocytes, transports O<sub>2</sub> to the tissues and returns CO<sub>2</sub> and protons to the lungs. Cyanide and carbon monoxide kill because they disrupt the physiologic function of the heme proteins cytochrome oxidase and hemoglobin, respectively. The secondary-tertiary structure of the subunits of hemoglobin resembles myoglobin. However, the tetrameric structure of hemoglobin permits cooperative interactions that are central to its function. For example, 2,3-BPG promotes the efficient release of O<sub>2</sub> by stabilizing the quaternary structure of deoxyhemoglobin. Hemoglobin and myoglobin illustrate both protein structure-function relationships and the molecular basis of genetic diseases such as sickle cell disease and the thalassemias.

## HEME & FERROUS IRON CONFER THE ABILITY TO STORE & TO TRANSPORT OXYGEN

Myoglobin and hemoglobin contain **heme**, a cyclic tetrapyrrole consisting of four molecules of pyrrole linked by methyne bridges. This planar network of conjugated double bonds absorbs visible light and colors heme deep red. The substituents at the  $\beta$ -positions of heme are methyl (M), vinyl (V), and propionate (Pr) groups arranged in the order M, V, M, V, M, Pr, Pr, M (**Figure 6–1**). The atom of ferrous iron ( $\text{Fe}^{2+}$ ) resides at the center of the planar tetrapyrrole. Other proteins with metal-containing tetrapyrrole prosthetic groups include the cytochromes (Fe and Cu) and chlorophyll (Mg) (see Chapter 31). Oxidation and reduction of the Fe and Cu atoms of cytochromes are essential to their biologic function as carriers of electrons. By contrast, oxidation of the  $\text{Fe}^{2+}$  of myoglobin or hemoglobin to  $\text{Fe}^{3+}$  destroys their biologic activity.

**Figure 6–1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

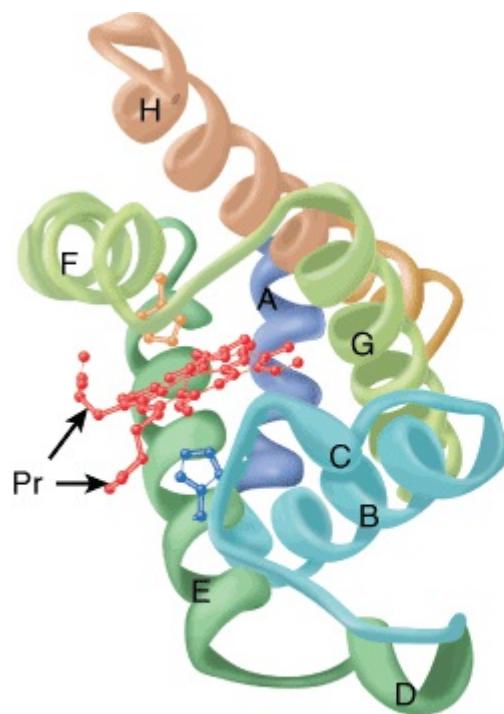
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Heme.** The pyrrole rings and methyne bridge carbons are coplanar, and the iron atom ( $\text{Fe}^{2+}$ ) resides in almost the same plane. The fifth and sixth coordination positions of  $\text{Fe}^{2+}$  are directly perpendicular to—and directly above and below—the plane of the heme ring. Observe the nature of the methyl (blue), vinyl (green), and propionate (orange) substituent groups on the  $\beta$ -carbons of the pyrrole rings, the central iron atom (red), and the location of the polar side of the heme ring (at about 7 o'clock) that faces the surface of the myoglobin molecule.

## Myoglobin Is Rich in $\alpha$ Helix

Oxygen stored in red muscle myoglobin is released during  $\text{O}_2$  deprivation (eg, severe exercise) for use in muscle mitochondria for aerobic synthesis of ATP (see Chapter 13). A 153-aminoacyl residue polypeptide (MW 17,000), myoglobin folds into a compact shape that measures  $4.5 \times 3.5 \times 2.5$  nm (**Figure 6–2**). An unusually high proportion, about 75%, of the residues are present in eight right-handed 7–20 residue  $\alpha$  helices. Starting at the amino terminal, these are termed helices A–H. Typical of globular proteins, the surface of myoglobin is rich in amino acids bearing polar and potentially charged side chains, while—with only two exceptions—the interior contains only residues such as Leu, Val, Phe, and Met that possess nonpolar R groups. The exceptions are His E7 and His F8, the seventh and eighth residues in helices E and F, which lie close to the heme iron, where they function in  $\text{O}_2$  binding.

**Figure 6–2**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Three-dimensional structure of myoglobin.** Shown is a ribbon diagram tracing the polypeptide backbone of myoglobin. The color of the polypeptide chain is graded along the visible spectrum from blue (N-terminal) to tan (C-terminal). The heme prosthetic group is red. The  $\alpha$ -helical regions are designated A through H. The distal (E7) and proximal (F8) histidine residues are highlighted in blue and orange, respectively. Note how the polar propionate substituents (Pr) project out of the heme toward solvent. (Adapted from Protein Data Bank ID no. 1a6n.)

## Histidines F8 & E7 Perform Unique Roles in Oxygen Binding

The heme of myoglobin lies in a crevice between helices E and F oriented with its polar propionate groups facing the surface of the globin (Figure 6–2). The remainder resides in the nonpolar interior. The fifth coordination position of the iron is occupied by a nitrogen from the imidazole ring of the **proximal histidine**, His F8. The **distal histidine**, His E7, lies on the side of the heme ring opposite to His F8.

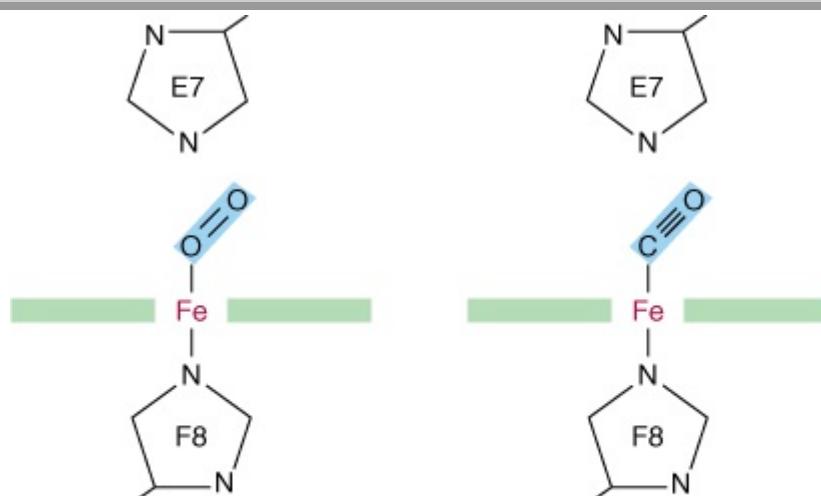
## The Iron Moves Toward the Plane of the Heme When Oxygen Is Bound

The iron of unoxygenated myoglobin lies 0.03 nm (0.3 Å) outside the plane of the heme ring, toward His F8. The heme therefore "puckers" slightly. When O<sub>2</sub> occupies the sixth coordination position, the iron moves to within 0.01 nm (0.1 Å) of the plane of the heme ring. Oxygenation of myoglobin thus is accompanied by motion of the iron, of His F8, and of residues linked to His F8.

## Apomyoglobin Provides a Hindered Environment for the Heme Iron

When O<sub>2</sub> binds to myoglobin, the bond between the first oxygen atom and the Fe<sup>2+</sup> is perpendicular to the plane of the heme ring. The bond that links the first and second oxygen atoms lies at an angle of 121° to the plane of the heme, orienting the second oxygen away from the distal histidine (Figure 6–3, left). This permits maximum overlap between the iron and one of the lone pairs of electrons on the sp<sup>2</sup> hybridized oxygen atoms, which lie at an angle of roughly 120° to the axis of the O=O double bond (Figure 6–4, left). Isolated heme binds carbon monoxide (CO) 25,000 times more strongly than oxygen. Since CO is present in small quantities in the atmosphere and arises in cells from the catabolism of heme, why is it that CO does not completely displace O<sub>2</sub> from heme iron? The accepted explanation is that the apoproteins of myoglobin and hemoglobin create a **hindered environment**. When CO binds to isolated heme, all the three atoms (Fe, C, and O) lie perpendicular to the plane of the heme. This geometry maximizes the overlap between the lone pair of electrons on the sp hybridized oxygen of the CO molecule and the Fe<sup>2+</sup> iron (Figure 6–4, right). However, in myoglobin and hemoglobin the distal histidine sterically precludes this preferred, high-affinity orientation of CO, but not that of O<sub>2</sub>. Binding at a less favored angle reduces the strength of the heme-CO bond to about 200 times that of the heme-O<sub>2</sub> bond (Figure 6–3, right) at which level the great excess of O<sub>2</sub> over CO normally present dominates. Nevertheless, about 1% of myoglobin typically is present combined with CO.

**Figure 6–3**

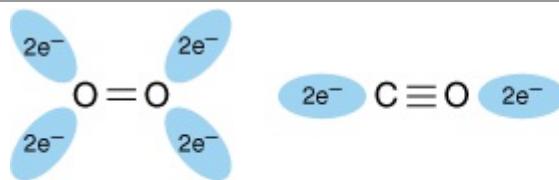


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Angles for bonding of oxygen and carbon monoxide (CO) to the heme iron of myoglobin.** The distal E7 histidine hinders bonding of CO at the preferred (90°) angle to the plane of the heme ring.

**Figure 6–4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

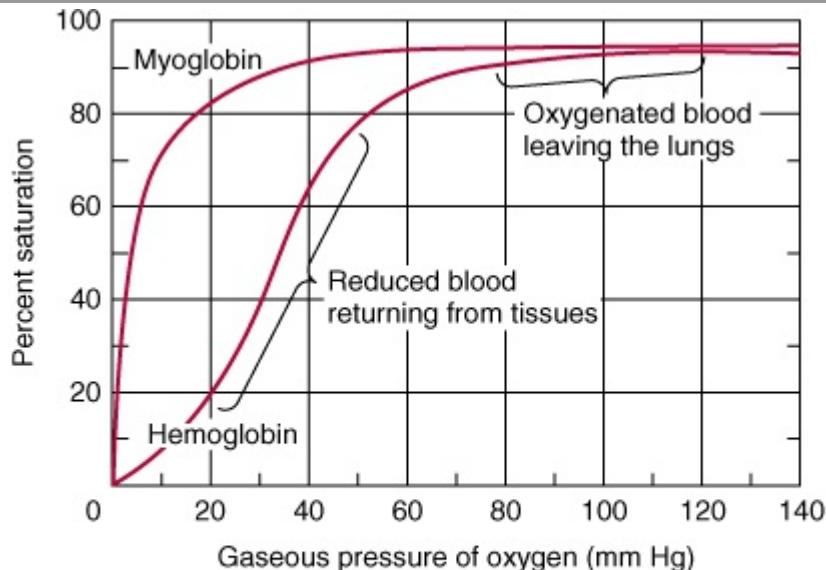
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Orientation of the lone pairs of electrons relative to the  $O=O$  and  $C\equiv O$  bonds of oxygen and carbon monoxide.** In molecular oxygen, formation of the double bond between the two oxygen atoms is facilitated by the adoption of an  $sp^2$  hybridization state by the valence electron of each oxygen atom. As a consequence, the two atoms of the oxygen molecule and each lone pair of electrons are coplanar and separated by an angle of roughly  $120^\circ$  (left). By contrast, the two atoms of carbon monoxide are joined by a triple bond, which requires that the carbon and oxygen atoms adopt an  $sp$  hybridization state. In this state the lone pairs of electrons and triple bonds are arranged in a linear fashion, where they are separated by an angle of  $180^\circ$  (right).

## THE OXYGEN DISSOCIATION CURVES FOR MYOGLOBIN & HEMOGLOBIN SUIT THEIR PHYSIOLOGIC ROLES

Why is myoglobin unsuitable as an O<sub>2</sub> transport protein but well suited for O<sub>2</sub> storage? The relationship between the concentration, or partial pressure, of O<sub>2</sub> (PO<sub>2</sub>) and the quantity of O<sub>2</sub> bound is expressed as an O<sub>2</sub> saturation isotherm (**Figure 6–5**). The oxygen-binding curve for myoglobin is hyperbolic. Myoglobin therefore loads O<sub>2</sub> readily at the PO<sub>2</sub> of the lung capillary bed (100 mm Hg). However, since myoglobin releases only a small fraction of its bound O<sub>2</sub> at the PO<sub>2</sub> values typically encountered in active muscle (20 mm Hg) or other tissues (40 mm Hg), it represents an ineffective vehicle for delivery of O<sub>2</sub>. When strenuous exercise lowers the PO<sub>2</sub> of muscle tissue to about 5 mm Hg, myoglobin releases O<sub>2</sub> for mitochondrial synthesis of ATP, permitting continued muscular activity.

**Figure 6–5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Oxygen-binding curves of both hemoglobin and myoglobin.** Arterial oxygen tension is about 100 mm Hg; mixed venous oxygen tension is about 40 mm Hg; capillary (active muscle) oxygen tension is about 20 mm Hg; and the minimum oxygen tension required for cytochrome oxidase is about 5 mm Hg. Association of chains into a tetrameric structure (hemoglobin) results in much greater oxygen delivery than would be possible with single chains. (Modified, with permission, from Scriver CR et al (editors): *The Molecular and Metabolic Bases of Inherited Disease*, 7th ed. McGraw-Hill, 1995.)

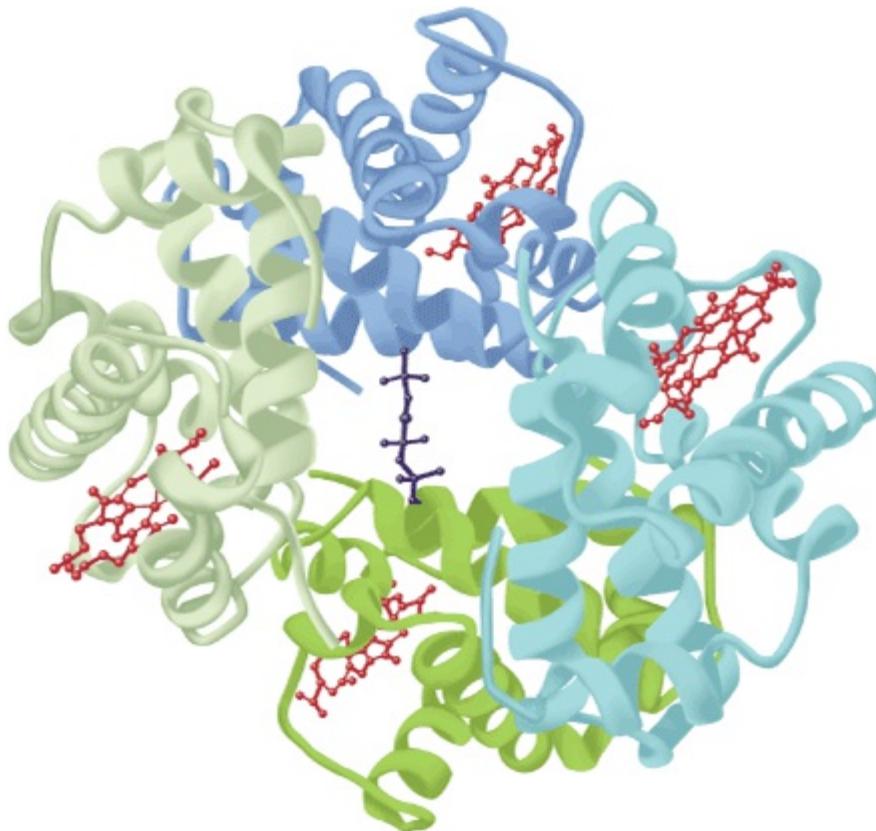
## THE ALLOSTERIC PROPERTIES OF HEMOGLOBINS RESULT FROM THEIR QUATERNARY STRUCTURES

The properties of individual hemoglobins are consequences of their quaternary as well as of their secondary and tertiary structures. The quaternary structure of hemoglobin confers striking additional properties, absent from monomeric myoglobin, which adapts it to its unique biologic roles. The **allosteric** (Gk *allos* "other," *stero*s "space") properties of hemoglobin provide, in addition, a model for understanding other allosteric proteins (see Chapter 18).

### Hemoglobin Is Tetrameric

Hemoglobins are tetramers composed of pairs of two different polypeptide subunits (Figure 6–6). Greek letters are used to designate each subunit type. The subunit composition of the principal hemoglobins are  $\alpha_2\beta_2$  (HbA; normal adult hemoglobin),  $\alpha_2\gamma_2$  (HbF; fetal hemoglobin),  $\alpha_2\delta_2$  (HbS; sickle cell hemoglobin), and  $\alpha_2\epsilon_2$  (HbA<sub>2</sub>; a minor adult hemoglobin). The primary structures of the  $\alpha$ ,  $\beta$ , and  $\gamma$  chains of human hemoglobin are highly conserved.

**Figure 6–6**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Hemoglobin.** Shown is the three-dimensional structure of deoxyhemoglobin with a molecule of 2,3-bisphosphoglycerate (dark blue) bound. The two  $\alpha$  subunits are colored in the darker shades of green and blue, the two  $\beta$  subunits in the lighter shades of green and blue, and the heme prosthetic groups in red. (Adapted from Protein Data Bank ID no. 1b86.)

### Myoglobin & the $\alpha$ Subunits of Hemoglobin Share Almost Identical Secondary and Tertiary Structures

Despite differences in the kind and number of amino acids present, myoglobin and the  $\alpha$  polypeptide of hemoglobin A have almost identical secondary and tertiary structures. Similarities include the location of the heme and the helical regions, and the presence of amino acids with similar properties at comparable locations. Although it possesses seven rather than eight helical regions, the  $\alpha$  polypeptide of hemoglobin also closely resembles myoglobin.

### Oxygenation of Hemoglobin Triggers Conformational Changes in the Apoprotein

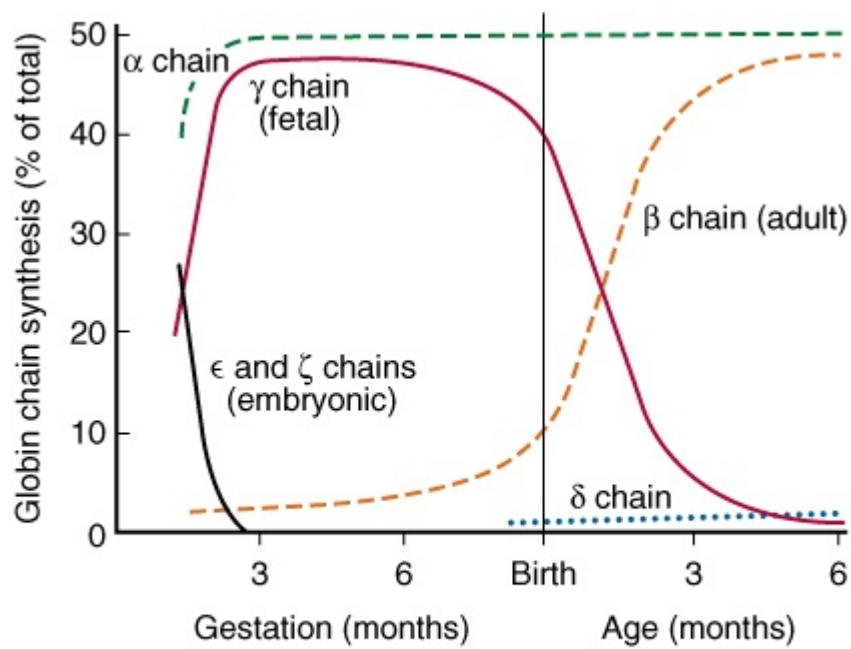
Hemoglobins bind four molecules of O<sub>2</sub> per tetramer, one per heme. A molecule of O<sub>2</sub> binds to a hemoglobin tetramer more readily if other O<sub>2</sub> molecules are already bound (Figure 6–5). Termed **cooperative binding**, this phenomenon permits hemoglobin to maximize both the quantity of O<sub>2</sub> loaded at the PO<sub>2</sub> of the lungs and the quantity of O<sub>2</sub> released at the PO<sub>2</sub> of the peripheral tissues. Cooperative interactions, an exclusive property of multimeric proteins, are critically important to aerobic life.

### P<sub>50</sub> Expresses the Relative Affinities of Different Hemoglobins for Oxygen

The quantity P<sub>50</sub>, a measure of O<sub>2</sub> concentration, is the partial pressure of O<sub>2</sub> that half-saturates a given hemoglobin. Depending on the organism, P<sub>50</sub> can vary widely, but in all instances, it will exceed the PO<sub>2</sub> of the peripheral tissues. For example, the values of P<sub>50</sub> for HbA and HbF are 26 and 20 mm Hg, respectively. In the placenta, this difference enables HbF to extract oxygen from the HbA in the mother's blood. However, HbF is suboptimal postpartum since its high affinity for O<sub>2</sub> limits the quantity of O<sub>2</sub> delivered to the tissues.

The subunit composition of hemoglobin tetramers undergoes complex changes during development. The human fetus initially synthesizes a  $\alpha_2\gamma_2$  tetramer. By the end of the first trimester,  $\alpha$  and  $\gamma$  subunits have been replaced by  $\alpha$  and  $\beta$  subunits, forming HbF ( $\alpha_2\beta_2$ ), the hemoglobin of late fetal life. While synthesis of  $\beta$  subunits begins in the third trimester,  $\beta$  subunits do not completely replace  $\gamma$  subunits to yield adult HbA ( $\alpha_2\beta_2$ ) until some weeks postpartum (Figure 6–7).

**Figure 6–7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

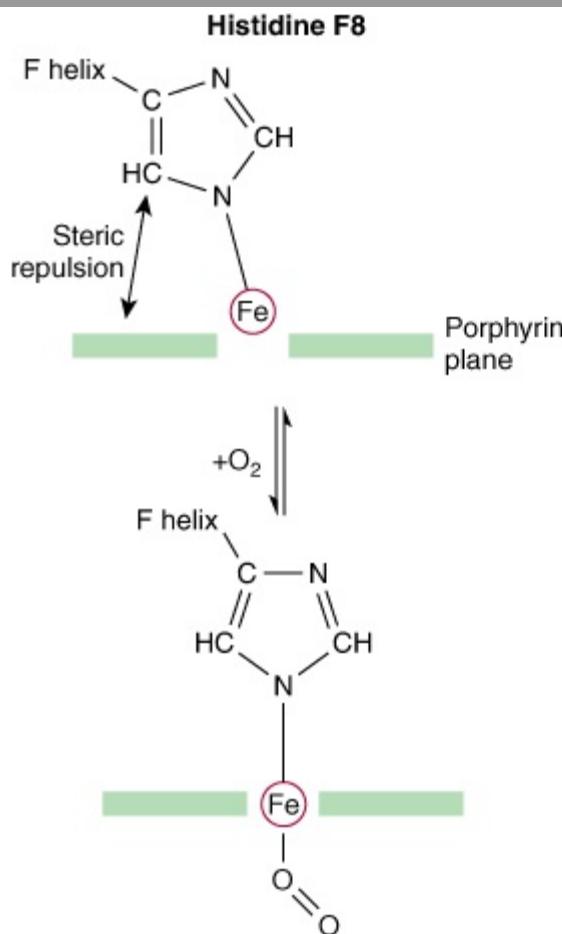
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Developmental pattern of the quaternary structure of fetal and newborn hemoglobins.** (Reproduced, with permission, from Ganong WF: *Review of Medical Physiology*, 20th ed. McGraw-Hill, 2001.)

### Oxygenation of Hemoglobin Is Accompanied by Large Conformational Changes

The binding of the first O<sub>2</sub> molecule to deoxyHb shifts the heme iron toward the plane of the heme ring from a position about 0.04 nm beyond it (Figure 6–8). This motion is transmitted to the proximal (F8) histidine and to the residues attached thereto, which in turn causes the rupture of salt bridges between the carboxyl terminal residues of all four subunits. As a result, one pair of  $\beta$  subunits rotates 15° with respect to the other, compacting the tetramer (Figure 6–9). Profound changes in secondary, tertiary, and quaternary structures accompany the O<sub>2</sub>-induced transition of hemoglobin from the low-affinity T (**taut state**) to the high-affinity R (**relaxed state**). These changes significantly increase the affinity of the remaining un oxygenated hemes for O<sub>2</sub>, as subsequent binding events require the rupture of fewer salt bridges (Figure 6–10). The terms T and R also are used to refer to the low-affinity and high-affinity conformations of allosteric enzymes, respectively.

**Figure 6–8**

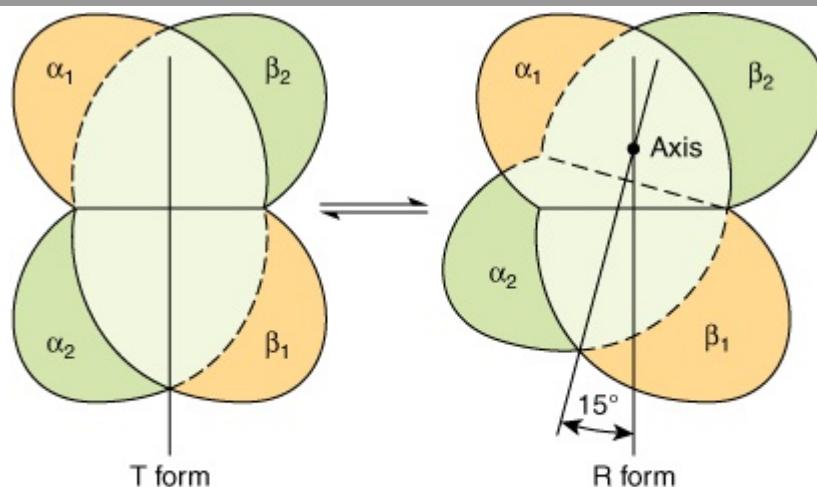


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**The iron atom moves into the plane of the heme on oxygenation.** Histidine F8 and its associated residues are pulled along with the iron atom. (Slightly modified and reproduced, with permission, from Stryer L: *Biochemistry*, 4th ed. Freeman, 1995. Copyright © 1995 W. H. Freeman and Company.)

**Figure 6–9**

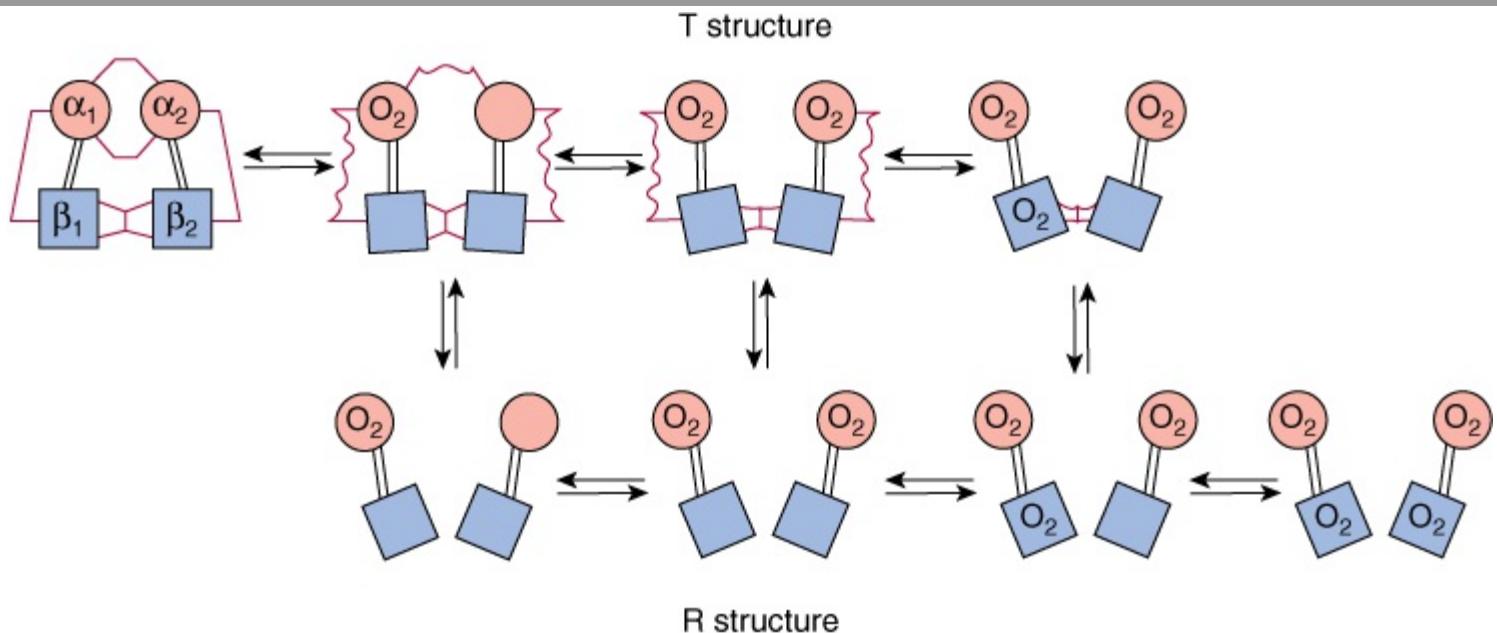


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

During transition of the T form to the R form of hemoglobin, the  $\alpha_2\beta_2$  pair of subunits (green) rotates through 15° relative to the pair of  $\alpha_1\beta_1$  subunits (yellow). The axis of rotation is eccentric, and the  $\alpha_2\beta_2$  pair also shifts toward the axis somewhat. In the representation, the tan  $\alpha_1\beta_1$  pair is shown fixed while the green  $\alpha_2\beta_2$  pair of subunits both shifts and rotates.

**Figure 6–10**



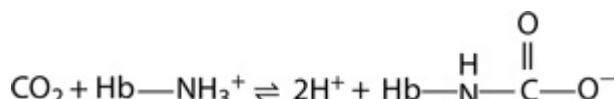
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Transition from the T structure to the R structure.** In this model, salt bridges (red lines) linking the subunits in the T structure break progressively as oxygen is added, and even those salt bridges that have not yet ruptured are progressively weakened (wavy red lines). The transition from T to R does not take place after a fixed number of oxygen molecules have been bound but becomes more probable as each successive oxygen binds. The transition between the two structures is influenced by protons, carbon dioxide, chloride, and BPG; the higher their concentration, the more oxygen must be bound to trigger the transition. Fully oxygenated molecules in the T structure and fully deoxygenated molecules in the R structure are not shown because they are unstable. (Modified and redrawn, with permission, from Perutz MF: Hemoglobin structure and respiratory transport. *Sci Am* [Dec] 1978; 239:92.)

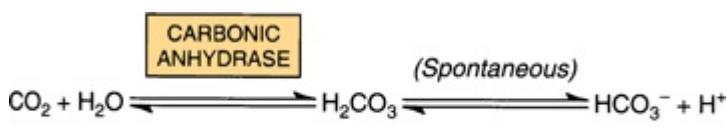
### After Releasing O<sub>2</sub> at the Tissues, Hemoglobin Transports CO<sub>2</sub> & Protons to the Lungs

In addition to transporting O<sub>2</sub> from the lungs to peripheral tissues, hemoglobin transports CO<sub>2</sub>, the byproduct of respiration, and protons from peripheral tissues to the lungs. Hemoglobin carries CO<sub>2</sub> as carbamates formed with the amino terminal nitrogens of the polypeptide chains:



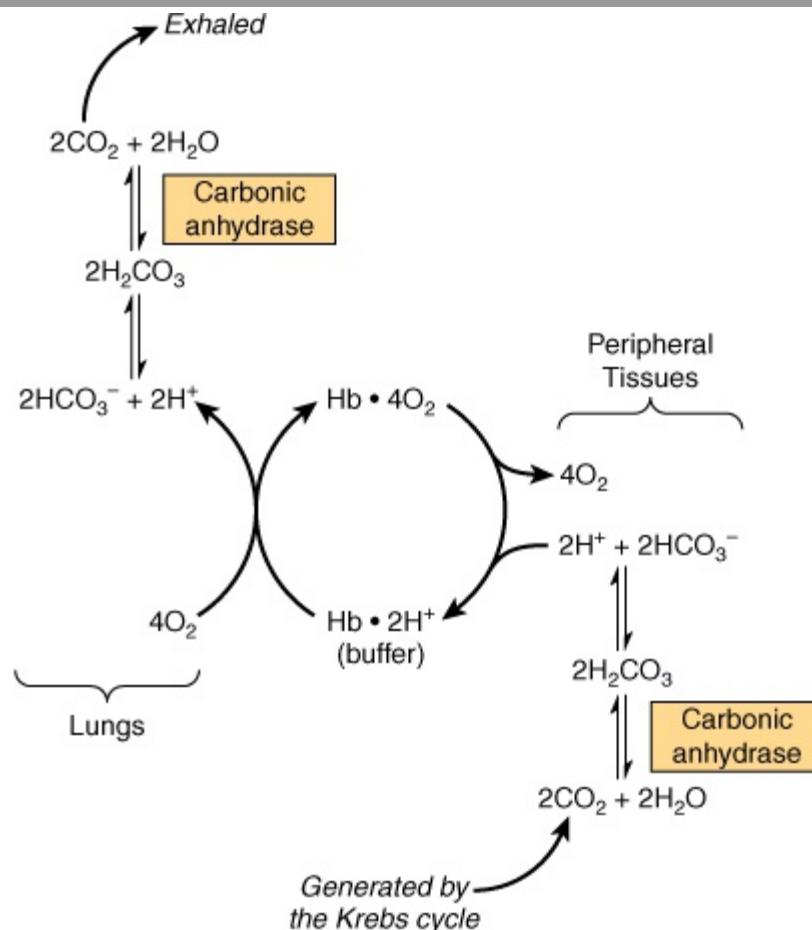
Carbamates change the charge on amino terminals from positive to negative, favoring salt bridge formation between  $\alpha$  and  $\beta$  chains.

Hemoglobin carbamates account for about 15% of the CO<sub>2</sub> in venous blood. Much of the remaining CO<sub>2</sub> is carried as bicarbonate, which is formed in erythrocytes by the hydration of CO<sub>2</sub> to carbonic acid (H<sub>2</sub>CO<sub>3</sub>), a process catalyzed by carbonic anhydrase. At the pH of venous blood, H<sub>2</sub>CO<sub>3</sub> dissociates into bicarbonate and a proton.



Deoxyhemoglobin binds one proton for every two O<sub>2</sub> molecules released, contributing significantly to the buffering capacity of blood. The somewhat lower pH of peripheral tissues, aided by carbamoylation, stabilizes the T state and thus enhances the delivery of O<sub>2</sub>. In lungs, the process reverses. As O<sub>2</sub> binds to deoxyhemoglobin, protons are released and combine with bicarbonate to form carbonic acid. Dehydration of H<sub>2</sub>CO<sub>3</sub>, catalyzed by carbonic anhydrase, forms CO<sub>2</sub>, which is exhaled. Binding of oxygen thus drives the exhalation of CO<sub>2</sub> (**Figure 6–11**). This reciprocal coupling of proton and O<sub>2</sub> binding is termed the **Bohr effect**. The Bohr effect is dependent upon **cooperative interactions between the hemes of the hemoglobin tetramer**. Myoglobin, a monomer, exhibits no Bohr effect.

**Figure 6–11**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

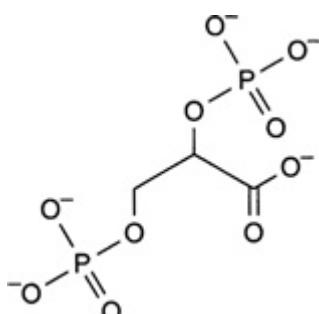
**The Bohr effect.** Carbon dioxide generated in peripheral tissues combines with water to form carbonic acid, which dissociates into protons and bicarbonate ions. Deoxyhemoglobin acts as a buffer by binding protons and delivering them to the lungs. In the lungs, the uptake of oxygen by hemoglobin releases protons that combine with bicarbonate ion, forming carbonic acid, which when dehydrated by carbonic anhydrase becomes carbon dioxide, which then is exhaled.

### Protons Arise from Rupture of Salt Bridges When O<sub>2</sub> Binds

Protons responsible for the Bohr effect arise from rupture of salt bridges during the binding of O<sub>2</sub> to T-state hemoglobin. Conversion to the oxygenated R state breaks salt bridges involving  $\alpha$ -chain residue His 146. The subsequent dissociation of protons from His 146 drives the conversion of bicarbonate to carbonic acid (**Figure 6–11**). Upon the release of O<sub>2</sub>, the T structure and its salt bridges re-form. This conformational change increases the pK<sub>a</sub> of the  $\alpha$ -chain His 146 residues, which bind protons. By facilitating the re-formation of salt bridges, an increase in proton concentration enhances the release of O<sub>2</sub> from oxygenated (R-state) hemoglobin. Conversely, an increase in PO<sub>2</sub> promotes proton release.

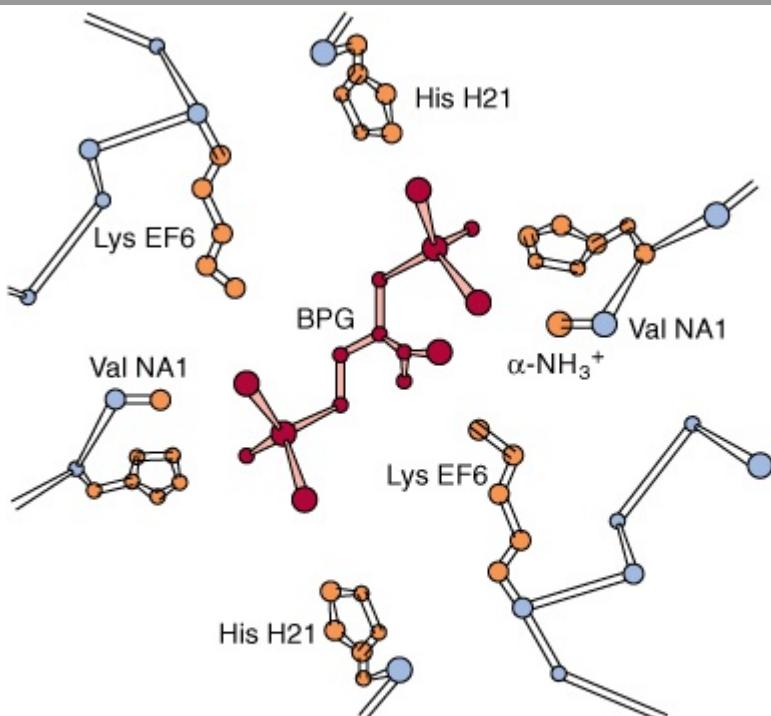
### 2,3-BPG Stabilizes the T Structure of Hemoglobin

A low PO<sub>2</sub> in peripheral tissues promotes the synthesis of 2,3-BPG in erythrocytes from the glycolytic intermediate 1,3-BPG.



The hemoglobin tetramer binds one molecule of BPG in the central cavity formed by its four subunits (**Figure 6–6**). However, the space between the  $\alpha$  helices of the  $\beta$  chains lining the cavity is sufficiently wide to accommodate BPG only when hemoglobin is in the T state. BPG forms salt bridges with the terminal amino groups of both  $\beta$  chains via Val N1A and with Lys E16 and His H21 (**Figure 6–12**). BPG therefore stabilizes deoxygenated (T-state) hemoglobin by forming additional salt bridges that must be broken prior to conversion to the R state.

**Figure 6–12**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Mode of binding of 2,3-bisphosphoglycerate (BPG) to human deoxyhemoglobin.** BPG interacts with three positively charged groups on each  $\beta$  chain. (Based on Arnone A: X-ray diffraction study of binding of 2,3-diphosphoglycerate to human deoxyhemoglobin. *Nature* 1972; 237:146. Copyright © 1972. Adapted by permission from Macmillan Publishers Ltd.)

Residue H21 of the  $\beta$  subunit of HbF is Ser rather than His. Since Ser cannot form a salt bridge, BPG binds more weakly to HbF than to HbA. The lower stabilization afforded to the T state by BPG accounts for HbF having a higher affinity for O<sub>2</sub> than HbA.

### Adaptation to High Altitude

Physiologic changes that accompany prolonged exposure to high altitude include an increase in the number of erythrocytes and in their concentrations of hemoglobin and of BPG. Elevated BPG lowers the affinity of HbA for O<sub>2</sub> (increases P<sub>50</sub>), which enhances the release of O<sub>2</sub> at peripheral tissues.

## NUMEROUS MUTATIONS AFFECTING HUMAN HEMOGLOBINS HAVE BEEN IDENTIFIED

Mutations in the genes that encode the  $\alpha$  or  $\beta$  subunits of hemoglobin potentially can affect its biologic function. However, almost all of the over 1,100 known genetic mutations affecting human hemoglobins are both extremely rare and benign, presenting no clinical abnormalities. When a mutation does compromise biologic function, the condition is termed a **hemoglobinopathy**. It is estimated that more than 7% of the globe's population are carriers for hemoglobin disorders. The URL <http://globin.cse.psu.edu/> (Globin Gene Server) provides information about—and links for—normal and mutant hemoglobins. Selected examples are described below.

### Methemoglobin & Hemoglobin M

In methemoglobinemia, the heme iron is ferric rather than ferrous. Methemoglobin thus can neither bind nor transport O<sub>2</sub>. Normally, the enzyme methemoglobin reductase reduces the Fe<sup>3+</sup> of methemoglobin to Fe<sup>2+</sup>. Methemoglobin can arise by oxidation of Fe<sup>2+</sup> to Fe<sup>3+</sup> as a side effect of agents such as sulfonamides, from hereditary hemoglobin M, or consequent to reduced activity of the enzyme methemoglobin reductase.

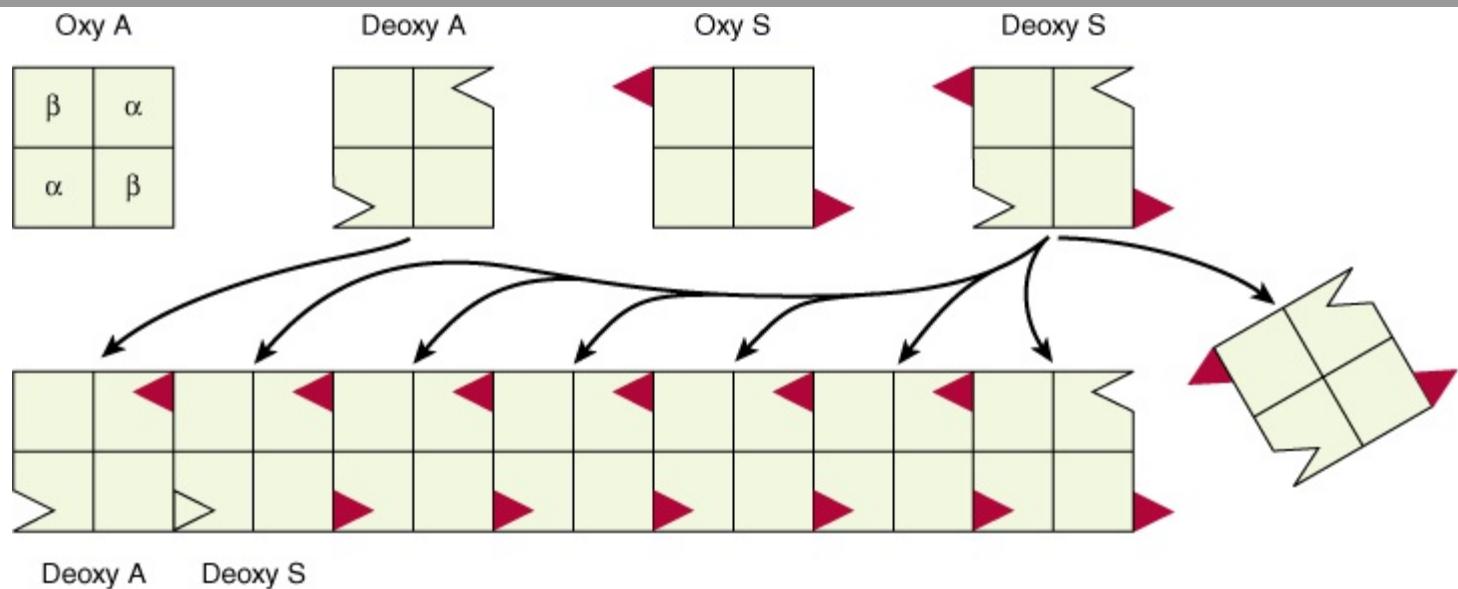
In hemoglobin M, histidine F8 (His F8) has been replaced by tyrosine. The iron of HbM forms a tight ionic complex with the phenolate anion of tyrosine that stabilizes the Fe<sup>3+</sup> form. In  $\alpha$ -chain hemoglobin M variants, the R-T equilibrium favors the T state. Oxygen affinity is reduced, and the Bohr effect is absent.  $\alpha$ -chain hemoglobin M variants exhibit R-T switching, and the Bohr effect is therefore present.

Mutations that favor the R state (eg, hemoglobin Chesapeake) increase O<sub>2</sub> affinity. These hemoglobins therefore fail to deliver adequate O<sub>2</sub> to peripheral tissues. The resulting tissue hypoxia leads to **polycythemia**, an increased concentration of erythrocytes.

### Hemoglobin S

In HbS, the nonpolar amino acid valine has replaced the polar surface residue Glu6 of the  $\beta$  subunit, generating a hydrophobic "sticky patch" on the surface of the  $\beta$  subunit of both oxyHbS and deoxyHbS. Both HbA and HbS contain a complementary sticky patch on their surfaces that is exposed only in the deoxygenated T state. Thus, at low PO<sub>2</sub>, deoxyHbS can polymerize to form long, insoluble fibers. Binding of deoxy-HbA terminates fiber polymerization, since HbA lacks the second sticky patch necessary to bind another Hb molecule (**Figure 6-13**). These twisted helical fibers distort the erythrocyte into a characteristic sickle shape, rendering it vulnerable to lysis in the interstices of the splenic sinusoids. They also cause multiple secondary clinical effects. A low PO<sub>2</sub>, such as that at high altitudes, exacerbates the tendency to polymerize. Emerging treatments for sickle cell disease include inducing HbF expression to inhibit the polymerization of HbS, stem cell transplantation, and, in the future, gene therapy.

**Figure 6-13**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Representation of the sticky patch (▲) on hemoglobin S and its "receptor" (▼) on deoxyhemoglobin A and deoxyhemoglobin S.** The complementary surfaces allow deoxyhemoglobin S to polymerize into a fibrous structure, but the presence of deoxyhemoglobin A will terminate the polymerization by failing to provide sticky patches. (Modified and reproduced, with permission, from Stryer L: *Biochemistry*, 4th ed. Freeman, 1995. Copyright © 1995 W. H. Freeman and Company.)

## BIOMEDICAL IMPLICATIONS

### Myoglobinuria

Following massive crush injury, myoglobin released from damaged muscle fibers colors the urine dark red. Myoglobin can be detected in plasma following a myocardial infarction, but assay of serum enzymes (see Chapter 7) provides a more sensitive index of myocardial injury.

### Anemias

Anemias, reductions in the number of red blood cells or of hemoglobin in the blood, can reflect impaired synthesis of hemoglobin (eg, in iron deficiency; Chapter 50) or impaired production of erythrocytes (eg, in folic acid or vitamin B12 deficiency; Chapter 44). Diagnosis of anemias begins with spectroscopic measurement of blood hemoglobin levels.

### Thalassemias

The genetic defects known as thalassemias result from the partial or total absence of one or more  $\alpha$  or  $\beta$  chains of hemoglobin. Over 750 different mutations have been identified, but only three are common. Either the  $\alpha$  chain (alpha thalassemias) or  $\beta$  chain (beta thalassemias) can be affected. A superscript indicates whether a subunit is completely absent ( $\alpha^0$  or  $\beta^0$ ) or whether its synthesis is reduced ( $\alpha^-$  or  $\beta^-$ ). Apart from marrow transplantation, treatment is symptomatic.

Certain mutant hemoglobins are common in many populations, and a patient may inherit more than one type. Hemoglobin disorders thus present a complex pattern of clinical phenotypes. The use of DNA probes for their diagnosis is considered in Chapter 39.

### Glycated Hemoglobin (HbA<sub>1c</sub>)

When blood glucose enters the erythrocytes, it glycates the  $\epsilon$ -amino group of lysyl residues and the amino terminals of hemoglobin. The fraction of hemoglobin glycated, normally about 5%, is proportionate to blood glucose concentration. Since the half-life of an erythrocyte is typically 60 days, the level of glycated hemoglobin (HbA<sub>1c</sub>) reflects the mean blood glucose concentration over the preceding 6–8 weeks. Measurement of HbA<sub>1c</sub> therefore provides valuable information for management of diabetes mellitus.

## SUMMARY

- Myoglobin is monomeric; hemoglobin is a tetramer of two subunit types ( $\alpha_2 \beta_2$  in HbA). Despite having different primary structures, myoglobin and the subunits of hemoglobin have nearly identical secondary and tertiary structures.
- Heme, an essentially planar, slightly puckered, cyclic tetrapyrrole has a central  $\text{Fe}^{2+}$  linked to all four nitrogen atoms of the heme, to histidine F8, and, in oxyMb and oxyHb, also to  $\text{O}_2$ .
- The  $\text{O}_2$ -binding curve for myoglobin is hyperbolic, but for hemoglobin it is sigmoidal, a consequence of cooperative interactions in the tetramer. Cooperativity maximizes the ability of hemoglobin both to load  $\text{O}_2$  at the  $\text{PO}_2$  of the lungs and to deliver  $\text{O}_2$  at the  $\text{PO}_2$  of the tissues.
- Relative affinities of different hemoglobins for oxygen are expressed as  $P_{50}$ , the  $\text{PO}_2$  that half-saturates them with  $\text{O}_2$ . Hemoglobins saturate at the partial pressures of their respective respiratory organ, eg, the lung or placenta.
- On oxygenation of hemoglobin, the iron, histidine F8, and linked residues move toward the heme ring. Conformational changes that accompany oxygenation include rupture of salt bonds and loosening of the quaternary structure, facilitating binding of additional  $\text{O}_2$ .
- 2,3-BPG in the central cavity of deoxyHb forms salt bonds with the  $\alpha$  subunits that stabilize deoxyHb. On oxygenation, the central cavity contracts, BPG is extruded, and the quaternary structure loosens.
- Hemoglobin also functions in  $\text{CO}_2$  and proton transport from tissues to lungs. Release of  $\text{O}_2$  from oxyHb at the tissues is accompanied by uptake of protons due to lowering of the  $pK_a$  of histidine residues.
- In sickle cell hemoglobin (HbS), Val replaces the  $\alpha_6$  Glu of HbA, creating a "sticky patch" that has a complement on deoxyHb (but not on oxyHb). DeoxyHbS polymerizes at low  $\text{O}_2$  concentrations, forming fibers that distort erythrocytes into sickle shapes.
- Alpha and beta thalassemias are anemias that result from reduced production of  $\alpha$  and  $\beta$  subunits of HbA, respectively.

## REFERENCES

- Frauenfelder H, McMahon BH, Fenimore PW: Myoglobin: The hydrogen atom of biology and a paradigm of complexity. *Proc Natl Acad Sci USA* 2003;100:8615.
- Hardison RC, Chui DH, Riemer C, et al: Databases of human hemoglobin variants and other resources at the globin gene server. *Hemoglobin* 2001;25:183.
- Lukin JA, Ho C: The structure-function relationship of hemoglobin in solution at atomic resolution. *Chem Rev* 2004;104:1219.
- Ordway GA, Garry DJ: Myoglobin: An essential hemoprotein in striated muscle. *J Exp Biol* 2004;207:3441.
- Papanikolaou E, Anagnos NP: Major challenges for gene therapy of thalassemia and sickle cell disease. *Curr Gene Ther* 2010;10:404.
- Schrier SL, Angelucci E: New strategies in the treatment of the thalassemias. *Annu Rev Med* 2005;56:157.
- Steinberg MH, Brugnara C: Pathophysiological-based approaches to treatment of sickle-cell disease. *Annu Rev Med* 2003;54:89.
- Umbreit J: Methemoglobin—it's not just blue: A concise review. *Am J Hematol* 2007;82:134.
- Weatherall DJ, Akinyanju O, Fucharoen S, et al: Inherited disorders of hemoglobin. In: *Disease Control Priorities in Developing Countries*, Jamison DT, Breman JG, Measham AR (editors). Oxford University Press and the World Bank, 2006;663–680.
- Weatherall DJ, Clegg JB, Higgs DR, et al: The hemoglobinopathies. In: *The Metabolic Basis of Inherited Disease*, 8th ed. Scriver CR, Sly WS, Childs B, et al (editors). McGraw-Hill, 2000;4571.
- Weatherall DJ, Clegg JD: *The Thalassemia Syndromes*. Blackwell Science, 2001.
- Yonetani T, Laberge M: Protein dynamics explain the allosteric behaviors of hemoglobin. *Biochim Biophys Acta* 2008;1784:1146.

## OBJECTIVES

After studying this chapter, you should be able to:

- Illustrate the structural relationships between B vitamins and coenzymes.
- Outline the four principal mechanisms by which enzymes achieve catalysis.
- Describe how an "induced fit" facilitates substrate recognition and catalysis.
- Outline the underlying principles of enzyme-linked immunoassays.
- Explain how coupling an enzyme to an NAD(P)<sup>+</sup>-dependent dehydrogenase can simplify assay of its activity.
- Identify enzymes and proteins whose plasma levels are used for the diagnosis and prognosis of a myocardial infarction.
- Describe the application of restriction endonucleases and of restriction fragment length polymorphisms in the detection of genetic diseases.
- Explain the utility of site-directed mutagenesis for the identification of residues involved in catalysis, in the recognition of substrates or allosteric effectors, or in the mechanism of enzyme action.
- Describe how the addition of fused affinity "tags" via recombinant DNA technology can facilitate purification of a protein expressed from its cloned gene.
- Indicate the function of specific proteases in the purification of affinity-tagged enzymes.
- Discuss the events that led to the discovery that RNAs can act as enzymes.

## BIOMEDICAL IMPORTANCE

Enzymes are biologic polymers that catalyze the chemical reactions that make life, as we know it, possible. The presence and maintenance of a complete and balanced set of enzymes is essential for the breakdown of nutrients to supply energy and chemical building blocks; the assembly of those building blocks into proteins, DNA, membranes, cells, and tissues; and the harnessing of energy to power cell motility, neural function, and muscle contraction. The vast majority of enzymes are proteins. Notable exceptions include **ribosomal RNAs** and a handful of self-cleaving or self-splicing RNA molecules known collectively as **ribozymes**. The ability to assay the activity of specific enzymes in blood, other tissue fluids, or cell extracts aids in the diagnosis and prognosis of disease. Deficiencies in the quantity or catalytic activity of key enzymes can result from genetic defects, nutritional deficits, or toxins. Defective enzymes can result from genetic mutations or infection by viral or bacterial pathogens (eg, *Vibrio cholerae*). Medical scientists address imbalances in enzyme activity by using pharmacologic agents to inhibit specific enzymes and are investigating gene therapy as a means to remedy deficits in enzyme level or function.

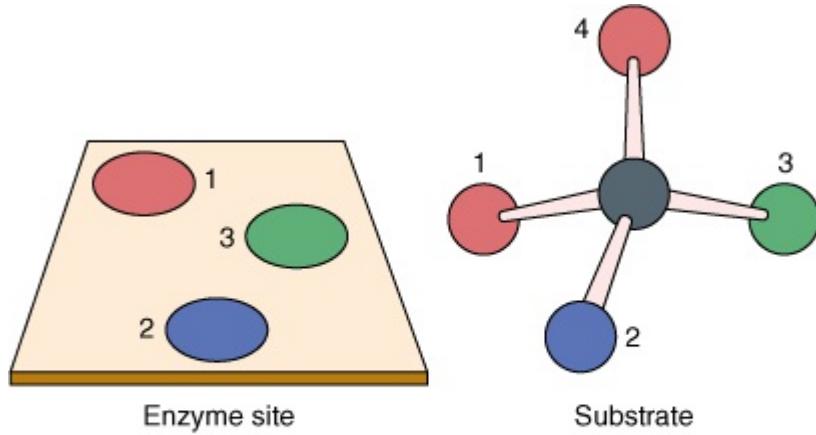
In addition to serving as the catalysts for all metabolic processes, their impressive catalytic activity, substrate specificity, and stereospecificity enable enzymes to fulfill key roles in other processes related to human health and well-being. The absolute stereospecificity of enzymes is of a particular value for use as soluble or immobilized catalysts for specific reactions in the synthesis of a drug or antibiotic. Proteases and amylases augment the capacity of detergents to remove dirt and stains. Enzymes play an important role in producing or enhancing the nutrient value of food products for both humans and animals. The protease rennin, for example, is utilized in the production of cheeses while lactase is employed to remove lactose from milk for the benefit of lactose-intolerant persons deficient in this hydrolytic enzyme (Chapter 43).

## ENZYMES ARE EFFECTIVE & HIGHLY SPECIFIC CATALYSTS

The enzymes that catalyze the conversion of one or more compounds (**substrates**) into one or more different compounds (**products**) enhance the rates of the corresponding noncatalyzed reaction by factors of at least  $10^6$ . Like all catalysts, enzymes are neither consumed nor permanently altered as a consequence of their participation in a reaction.

In addition to being highly efficient, enzymes are also extremely selective catalysts. Unlike most catalysts used in synthetic chemistry, enzymes are specific both for the type of reaction catalyzed and for a single substrate or a small set of closely related substrates. Enzymes are also stereospecific catalysts and typically catalyze reactions of only one stereoisomer of a given compound—for example, D- but not L-sugars, L- but not D-amino acids. Since they bind substrates through at least "three points of attachment," enzymes can even convert non-chiral substrates to chiral products. **Figure 7-1** illustrates why the enzyme-catalyzed reduction of the nonchiral substrate pyruvate produces only L-lactate, not a racemic mixture of D- and L-lactate. The exquisite specificity of enzyme catalysts imbues living cells with the ability to simultaneously conduct and independently control a broad spectrum of chemical processes.

**Figure 7-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Planar representation of the "three-point attachment" of a substrate to the active site of an enzyme.** Although atoms 1 and 4 are identical, once atoms 2 and 3 are bound to their complementary sites on the enzyme, only atom 1 can bind. Once bound to an enzyme, apparently identical atoms thus may be distinguishable, permitting a stereospecific chemical change.

## ENZYMES ARE CLASSIFIED BY REACTION TYPE

The commonly used names for most enzymes describe the type of reaction catalyzed, followed by the suffix **-ase**. For example, dehydrogenases remove hydrogen atoms, proteases hydrolyze proteins, and isomerases catalyze rearrangements in configuration. Modifiers may precede the name to indicate the substrate (*xanthine oxidase*), the source of the enzyme (*pancreatic ribonuclease*), its regulation (*hormone-sensitive lipase*), or a feature of its mechanism of action (*cysteine protease*). Where needed, alphanumeric designators are added to identify multiple forms of an enzyme (eg, RNA polymerase *III*; protein kinase *C* ).

To address ambiguities, the International Union of Biochemists (IUB) developed an unambiguous system of enzyme nomenclature in which each enzyme has a unique name and code number that identify the type of reaction catalyzed and the substrates involved. Enzymes are grouped into the following six classes.

1. **Oxidoreductases**—enzymes that catalyze oxidations and reductions.
2. **Transferases**—enzymes that catalyze transfer of moieties such as glycosyl, methyl, or phosphoryl groups.
3. **Hydrolases**—enzymes that catalyze *hydrolytic* cleavage of C—C, C—O, C—N and other covalent bonds.
4. **Lyases**—enzymes that catalyze cleavage of C—C, C—O, C—N and other covalent bonds by *atom elimination*, generating double bonds.
5. **Isomerases**—enzymes that catalyze geometric or structural changes *within* a molecule.
6. **Ligases**—enzymes that catalyze the joining together (ligation) of two molecules in reactions coupled to the hydrolysis of ATP.

Despite the clarity of the IUB system, the names are lengthy and relatively cumbersome, so we generally continue to refer to enzymes by their traditional, albeit sometimes ambiguous names. The IUB name for hexokinase illustrates both the clarity of the IUB system and its complexities. The IUB name of hexokinase is ATP:D-hexose 6-phosphotransferase E.C. 2.7.1.1. This name identifies hexokinase as a member of class 2 (transferases), subclass 7 (transfer of a phosphoryl group), subsubclass 1 (alcohol is the phosphoryl acceptor), and "hexose-6" indicates that the alcohol phosphorylated is on carbon six of a hexose. However, we continue to call it hexokinase.

## PROSTHETIC GROUPS, COFACTORS, & COENZYMES PLAY IMPORTANT ROLES IN CATALYSIS

Many enzymes contain small nonprotein molecules and metal ions that participate directly in substrate binding or in catalysis. Termed **prosthetic groups, cofactors, and coenzymes**, these extend the repertoire of catalytic capabilities beyond those afforded by the limited number of functional groups present on the aminoacyl side chains of peptides.

### Prosthetic Groups Are Tightly Integrated into an Enzyme's Structure

Prosthetic groups are tightly and stably incorporated into a protein's structure by covalent or noncovalent forces. Examples include pyridoxal phosphate, flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), thiamin pyrophosphate, biotin, and the metal ions of Co, Cu, Mg, Mn, and Zn. Metals are the most common prosthetic groups. The roughly one-third of all enzymes that contain tightly bound metal ions are termed **metalloenzymes**. Metal ions that participate in redox reactions generally are complexed to prosthetic groups such as heme (Chapter 6) or iron-sulfur clusters (Chapter 12). Metals also may facilitate the binding and orientation of substrates, the formation of covalent bonds with reaction intermediates ( $\text{Co}^{2+}$  in coenzyme  $\text{B}_{12}$ ), or by acting as Lewis acids or bases to render substrates more **electrophilic** (electron-poor) or **nucleophilic** (electron-rich), and hence more reactive.

### Cofactors Associate Reversibly with Enzymes or Substrates

**Cofactors** serve functions similar to those of prosthetic groups, but bind in a transient, dissociable manner either to the enzyme or to a substrate such as ATP. Unlike the stably associated prosthetic groups, cofactors must be present in the medium surrounding the enzyme for catalysis to occur. The most common cofactors also are metal ions. Enzymes that require a metal ion cofactor are termed **metal-activated enzymes** to distinguish them from the **metalloenzymes** for which metal ions serve as prosthetic groups.

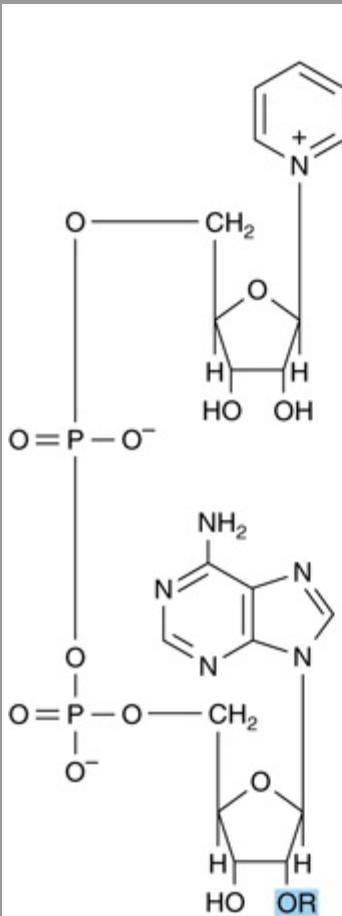
### Coenzymes Serve as Substrate Shuttles

**Coenzymes** serve as recyclable shuttles—or group transfer agents—that transport many substrates from one point within the cell to another. The function of these shuttles is twofold. First, they stabilize species such as hydrogen atoms ( $\text{FADH}_2$ ) or hydride ions ( $\text{NADH}$ ) that are too reactive to persist for any significant time period in the presence of the water or organic molecules that permeate the cell interior. They also serve as an adaptor or handle that facilitates the recognition and binding of small chemical groups, such as acetate (coenzyme A), by their target enzymes. Other chemical moieties transported by coenzymes include methyl groups (folates) and oligosaccharides (dolichol).

### Many Coenzymes, Cofactors & Prosthetic Groups Are Derivatives of B Vitamins

The water-soluble B vitamins supply important components of numerous coenzymes. Several coenzymes contain, in addition, the adenine, ribose, and phosphoryl moieties of AMP or ADP (Figure 7–2). **Nicotinamide** is a component of the redox coenzymes NAD and NADP, whereas **riboflavin** is a component of the redox coenzymes FMN and FAD. **Pantothenic acid** is a component of the acyl group carrier **coenzyme A**. As its pyrophosphate, **thiamin** participates in decarboxylation of  $\alpha$ -keto acids, and the **folic acid** and **cobamide** coenzymes function in one-carbon metabolism.

Figure 7–2



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

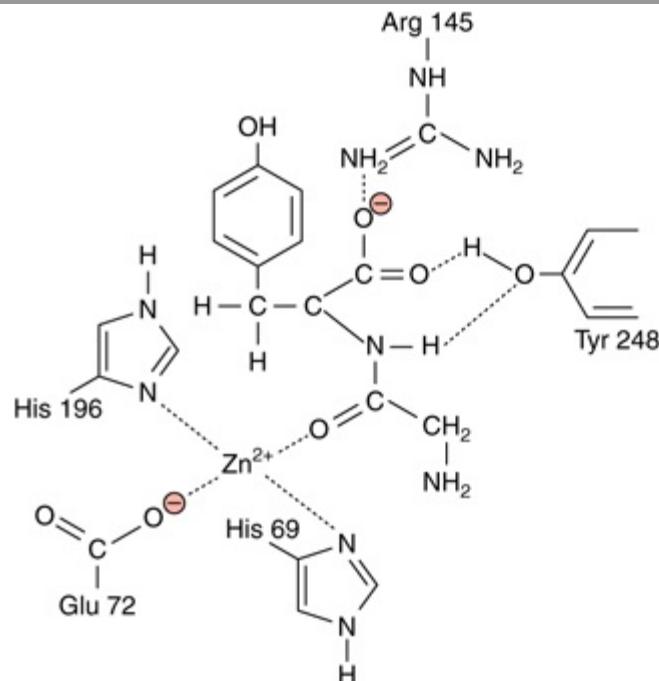
Structure of  $\text{NAD}^+$  and  $\text{NADP}^+$ . For  $\text{NAD}^+$ , R=H. For  $\text{NADP}^+$ , R= $\text{PO}_3^{2-}$ .

## CATALYSIS OCCURS AT THE ACTIVE SITE

An important early 20th-century insight into enzymic catalysis sprang from the observation that the presence of substrates renders enzymes more resistant to the denaturing effects of elevated temperatures. This observation led Emil Fischer to propose that enzymes and their substrates interact to form an enzyme–substrate (ES) complex whose thermal stability was greater than that of the enzyme itself. This insight profoundly shaped our understanding of both the chemical nature and kinetic behavior (Chapter 8) of enzymic catalysis.

Fischer reasoned that the exquisitely high specificity with which enzymes discriminate their substrates when forming an ES complex was analogous to the manner in which a mechanical lock distinguishes the proper key. In most enzymes, the "lock" is formed by a cleft or pocket on the protein's surface that forms part of a region called the **active site** (**Figures 5–6** and **5–8**). As implied by the adjective "active," the active site is much more than simply a recognition site for binding substrates. Within the active site, substrates are brought into close proximity to one another in optimal alignment with the cofactors, prosthetic groups, and amino acid side chains responsible for catalyzing their chemical transformation into products (**Figure 7–3**). Catalysis is further enhanced by the capacity of the active site to shield substrates from water and generate an environment whose polarity, hydrophobicity, acidity, or alkalinity can differ markedly from that of the surrounding cytoplasm.

**Figure 7–3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two-dimensional representation of a dipeptide substrate, glycyl-tyrosine, bound within the active site of carboxypeptidase A.

## ENZYMES EMPLOY MULTIPLE MECHANISMS TO FACILITATE CATALYSIS

Enzymes use various combinations of four general mechanisms to achieve dramatic catalytic enhancement of the rates of chemical reactions.

### Catalysis by Proximity

For molecules to react, they must come within bond-forming distance of one another. The higher their concentration, the more frequently they will encounter one another, and the greater will be the rate of their reaction. When an enzyme binds substrate molecules at its active site, it creates a region of high local substrate concentration. This environment also orients the substrate molecules spatially in a position ideal for them to interact, resulting in rate enhancements of at least a thousandfold.

### Acid-Base Catalysis

The ionizable functional groups of aminoacyl side chains and (where present) of prosthetic groups contribute to catalysis by acting as acids or bases. Acid-base catalysis can be either specific or general. By "specific" we mean only protons ( $\text{H}_3\text{O}^+$ ) or  $\text{OH}^-$  ions. In **specific acid catalysis** or **specific base catalysis**, the rate of reaction is sensitive to changes in the concentration of protons or of *independent* of the concentrations of other acids (proton donors) or bases (proton acceptors) present in the solution or at the active site. Reactions whose rates are responsive to *all* the acids or bases present are said to be subject to **general acid catalysis** or **general base catalysis**.

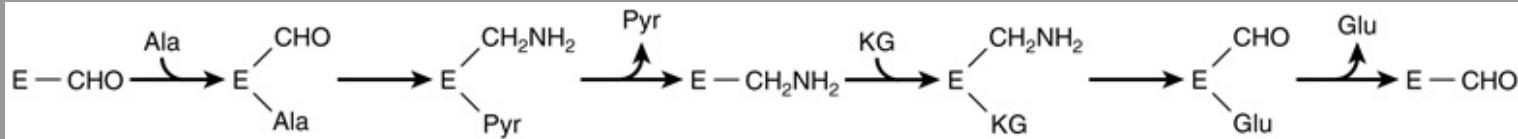
### Catalysis by Strain

Enzymes that catalyze *lytic* reactions that involve breaking a covalent bond typically bind their substrates in a conformation that is somewhat unfavorable for the bond that will undergo cleavage. This conformation mimics that of the **transition state intermediate**, a transient species that represents the transition state, or half-way point, in the transformation of substrates to products. The resulting strain stretches or distorts the targeted bond, weakening it and making it more vulnerable to cleavage. Nobel Laureate Linus Pauling was the first to suggest a role for transition state stabilization as a general mechanism by which enzymes accelerate the rates of chemical reactions. Knowledge of the transition state of an enzyme-catalyzed reaction is frequently exploited by chemists to design and create more effective enzyme inhibitors, called **transition state analogs**, as potential pharmacophores.

### Covalent Catalysis

The process of **covalent catalysis** involves the formation of a covalent bond between the enzyme and one or more substrates. The modified enzyme then becomes a reactant. Covalent catalysis introduces a new reaction pathway whose activation energy is lower—and the reaction therefore is faster—than the reaction pathway in homogeneous solution. Chemical modification of the enzyme is, however, transient. Completion of the reaction returns the enzyme to its original unmodified state. Its role thus remains catalytic. Covalent catalysis is particularly common among enzymes that catalyze group transfer reactions. Residues on the enzyme that participate in covalent catalysis generally are cysteine or serine and occasionally histidine. Covalent catalysis often follows a "ping-pong" mechanism—one in which the first substrate is bound and its product released prior to the binding of the second substrate (Figure 7-4).

**Figure 7-4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

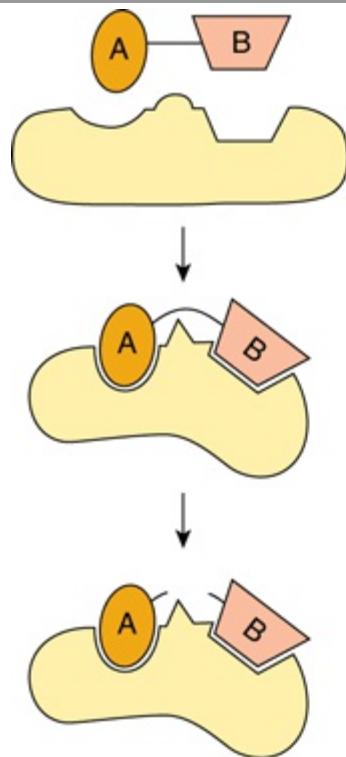
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Ping-pong" mechanism for transamination. E-CHO and E-CH<sub>2</sub>NH<sub>2</sub> represent the enzyme-pyridoxal phosphate and enzyme-pyridoxamine complexes, respectively. (Ala, alanine; Glu, glutamate; KG, α-ketoglutarate; Pyr, pyruvate.)

## SUBSTRATES INDUCE CONFORMATIONAL CHANGES IN ENZYMES

While Fischer's "lock and key model" accounted for the exquisite specificity of enzyme-substrate interactions, the implied rigidity of the enzyme's active site failed to account for the dynamic changes that we now know accompany catalysis. This drawback was addressed by Daniel Koshland's **induced fit** model, which states that when substrates approach and bind to an enzyme they induce a conformational change analogous to placing a hand (substrate) into a glove (enzyme) (Figure 7–5). The enzyme in turn induces reciprocal changes in its substrates, harnessing the energy of binding to facilitate the transformation of substrates into products. The induced fit model has been amply confirmed by biophysical studies of enzyme motion during substrate binding.

**Figure 7–5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

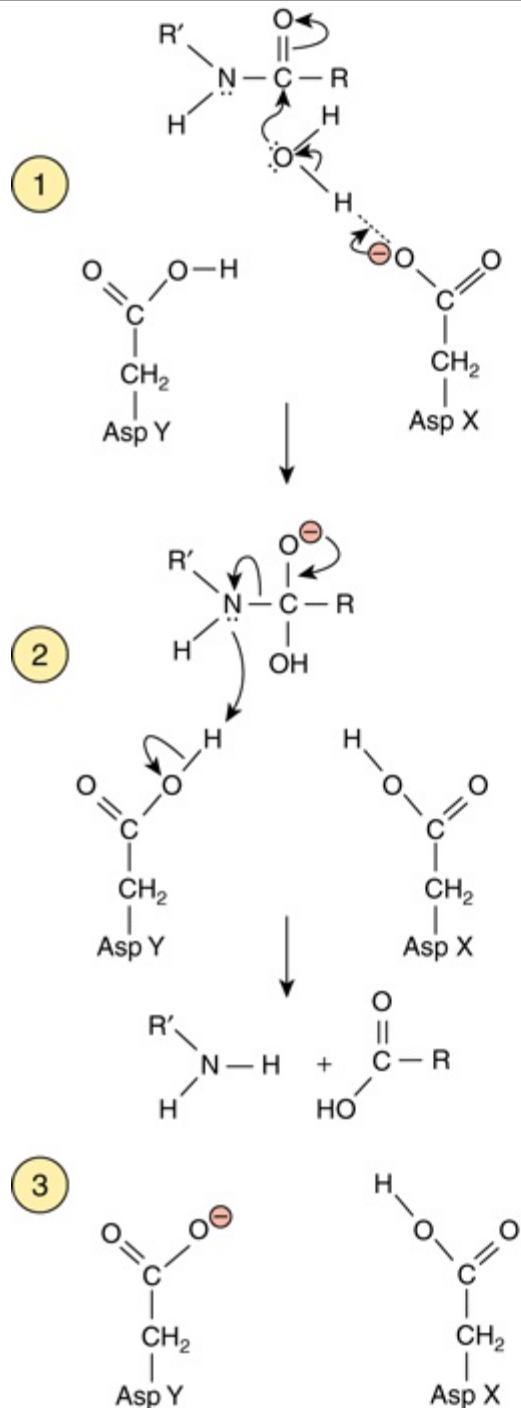
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Two-dimensional representation of Koshland's induced fit model of the active site of a lyase.** Binding of the substrate A–B induces conformational changes in the enzyme that align catalytic residues which participate in catalysis and strains the bond between A and B, facilitating its cleavage.

## HIV PROTEASE ILLUSTRATES ACID-BASE CATALYSIS

Enzymes of the **aspartic protease family**, which includes the digestive enzyme pepsin, the lysosomal cathepsins, and the protease produced by the human immunodeficiency virus (HIV) share a common catalytic mechanism. Catalysis involves two conserved aspartyl residues, which act as acid-base catalysts. In the first stage of the reaction, an aspartate functioning as a general base (Asp X, **Figure 7–6**) extracts a proton from a water molecule, making it more nucleophilic. The resulting nucleophile then attacks the electrophilic carbonyl carbon of the peptide bond targeted for hydrolysis, forming a **tetrahedral transition state intermediate**. A second aspartate (Asp Y, **Figure 7–6**) then facilitates the decomposition of this tetrahedral intermediate by donating a proton to the amino group produced by rupture of the peptide bond. The two different active site aspartates can act simultaneously as a general base or as a general acid because their immediate environment favors ionization of one, but not the other.

**Figure 7–6**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Mechanism for catalysis by an aspartic protease such as HIV protease.** Curved arrows indicate directions of electron movement. ① Aspartate X acts as a base to activate a water molecule by abstracting a proton. ② The activated water molecule attacks the peptide bond, forming a transient tetrahedral intermediate. ③ Aspartate Y acts as an acid to facilitate breakdown of the tetrahedral intermediate and release of the split products by donating a proton to the newly formed amino group. Subsequent shuttling of the proton on Asp X to Asp Y restores the protease to its initial state.

## CHYMOTRYPSIN & FRUCTOSE-2, 6-BISPHOSPHATASE ILLUSTRATE COVALENT CATALYSIS

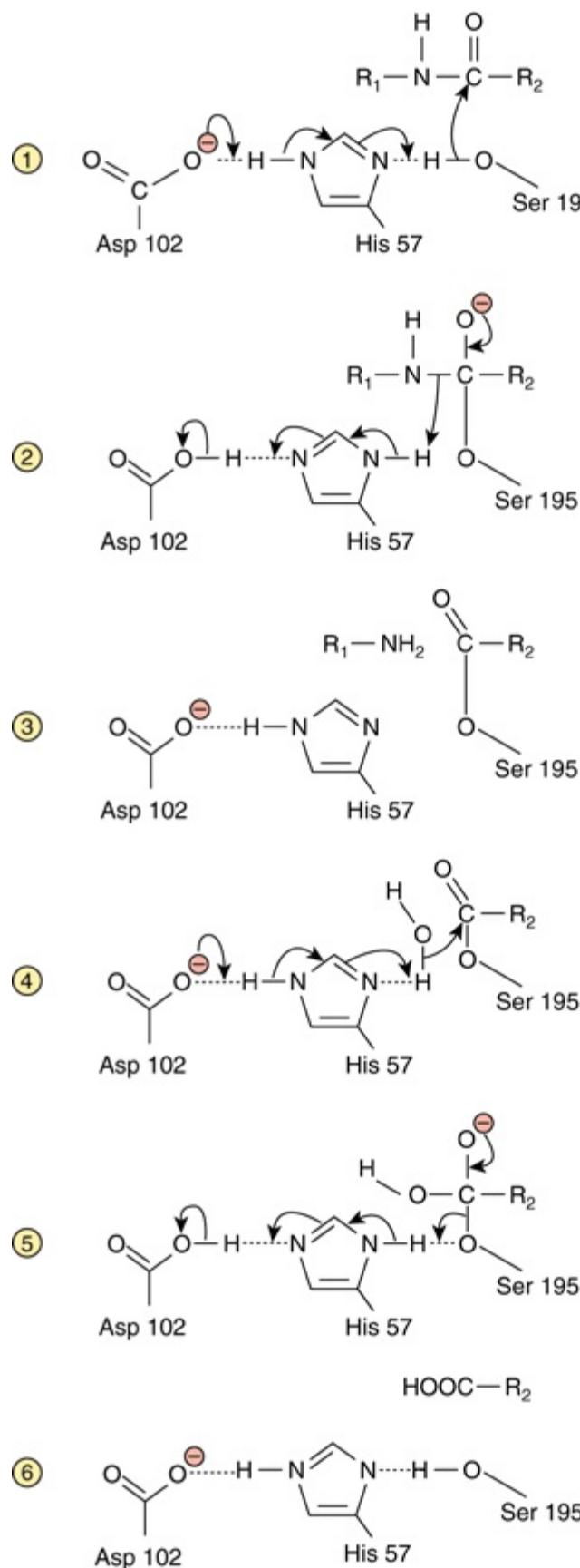
### Chymotrypsin

While catalysis by aspartic proteases involves the direct hydrolytic attack of water on a peptide bond, catalysis by the **serine protease** chymotrypsin involves prior formation of a covalent acyl-enzyme intermediate. A highly reactive seryl residue, serine 195, participates in a charge-relay network with histidine 57 and aspartate 102. While these three residues are far apart in primary structure, in the active site of the mature, folded protein they are within bond-forming distance of one another. Aligned in the order Asp 102-His 57-Ser 195, they constitute a "charge-relay network" that functions as a "**proton shuttle**."

Binding of substrate initiates proton shifts that in effect transfer the hydroxyl proton of Ser 195 to Asp 102 (**Figure 7–7**). The enhanced nucleophilicity of the seryl oxygen facilitates its attack on the carbonyl carbon of the peptide bond of the substrate, forming a covalent **acyl-enzyme intermediate**. The proton on Asp 102 then shuttles via His 57 to the amino group liberated when the peptide bond is cleaved. The portion of the original peptide with a free amino group then leaves the active site and is replaced by a water molecule. The charge-relay network now activates the water molecule by withdrawing a proton through His 57 to Asp 102. The resulting hydroxide ion attacks the acyl-enzyme intermediate and a reverse proton shuttle returns a proton to Ser 195, restoring its original state. While modified during the process of catalysis, chymotrypsin emerges unchanged on completion of the reaction. The protease trypsin and elastase employ a similar catalytic mechanism, but the numbers of the residues in their Ser-His-Asp proton shuttles differ.

**Figure 7–7**





Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

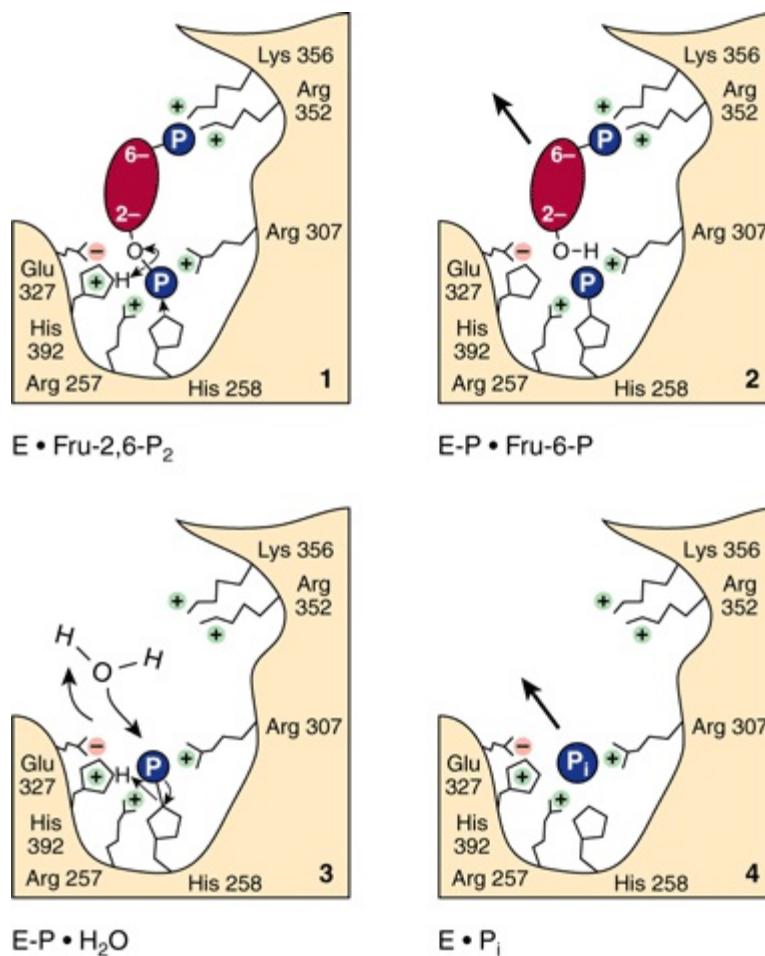
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Catalysis by chymotrypsin.** ① The charge-relay system removes a proton from Ser 195, making it a stronger nucleophile. ② Activated Ser 195 attacks the peptide bond, forming a transient tetrahedral intermediate. ③ Release of the amino terminal peptide is facilitated by donation of a proton to the newly formed amino group by His 57 of the charge-relay system, yielding an acyl-Ser 195 intermediate. ④ His 57 and Asp 102 collaborate to activate a water molecule, which attacks the acyl-Ser 195, forming a second tetrahedral intermediate. ⑤ The charge-relay system donates a proton to Ser 195, facilitating breakdown of tetrahedral intermediate to release the carboxyl terminal peptide⑥.

### Fructose-2,6-Bisphosphatase

Fructose-2,6-bisphosphatase, a regulatory enzyme of gluconeogenesis (Chapter 20), catalyzes the hydrolytic release of the phosphate on carbon 2 of fructose 2,6-bisphosphate. Figure 7–8 illustrates the roles of seven active site residues. Catalysis involves a "catalytic triad" of one Glu and two His residues and a covalent phosphohistidyl intermediate.

Figure 7–8



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Catalysis by fructose-2,6-bisphosphatase.** (1) Lys 356 and Arg 257, 307, and 352 stabilize the quadruple negative charge of the substrate by charge-charge interactions. Glu 327 stabilizes the positive charge on His 392. (2) The nucleophile His 392 attacks the C-2 phosphoryl group and transfers it to His 258, forming a phosphoryl-enzyme intermediate. Fructose 6-phosphate now leaves the enzyme. (3) Nucleophilic attack by a water molecule, possibly assisted by Glu 327 acting as a base, forms inorganic phosphate. (4) Inorganic orthophosphate is released from Arg 257 and Arg 307. (Reproduced, with permission, from Pilkis SJ, et al: 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: A metabolic signaling enzyme. *Annu Rev Biochem* 1995;64:799. © 1995 by Annual Reviews, [www.annualreviews.org](http://www.annualreviews.org).)

## Catalytic Residues Are Highly Conserved

Members of an enzyme family such as the aspartic or serine proteases employ a similar mechanism to catalyze a common reaction type, but act on different substrates. Most enzyme families arose through gene duplication events that created a second copy of the gene that encodes a particular enzyme. The proteins encoded by the two genes can then evolve independently to recognize different substrates—resulting, for example, in chymotrypsin, which cleaves peptide bonds on the carboxyl terminal side of large hydrophobic amino acids, and trypsin, which cleaves peptide bonds on the carboxyl terminal side of basic amino acids. Proteins that diverged from a common ancestor are said to be **homologous** to one another. The common ancestry of enzymes can be inferred from the presence of specific amino acids in the same position in each family member. These residues are said to be **conserved residues**. Table 7–1 illustrates the primary structural conservation of two components of the charge-relay network for several serine proteases. Among the most highly conserved residues are those that participate directly in catalysis.

**Table 7–1 Amino Acid Sequences in the Neighborhood of the Catalytic Sites of Several Bovine Proteases**

Enzyme	Sequence Around Serine (S)	Sequence Around Histidine (H)
Trypsin	D S C Q D G (S) G G P V V C S G	K V V S A A (H) C Y K S G
Chymotrypsin A	S S C M G D (S) G G P L V C K K	N V V T A A (H) G G V T T
Chymotrypsin B	S S C M G D (S) G G P L V C Q K	N V V T A A (H) C G V T T
Thrombin	D A C E G D (S) G G P F V M K S	P V L T A A (H) C L L Y P

**Note:** Regions shown are those on either side of the catalytic site seryl (S) and histidyl (H) residues.

## ISOZYMES ARE DISTINCT ENZYME FORMS THAT CATALYZE THE SAME REACTION

Higher organisms often elaborate several physically distinct versions of a given enzyme, each of which catalyzes the same reaction. Like the members of other protein families, these protein catalysts or **isozymes** arise through gene duplication. Isozymes may exhibit subtle differences in properties such as sensitivity to particular regulatory factors (Chapter 9) or substrate affinity (eg, hexokinase and glucokinase) that adapt them to specific tissues or circumstances. Some isozymes may also enhance survival by providing a "backup" copy of an essential enzyme.

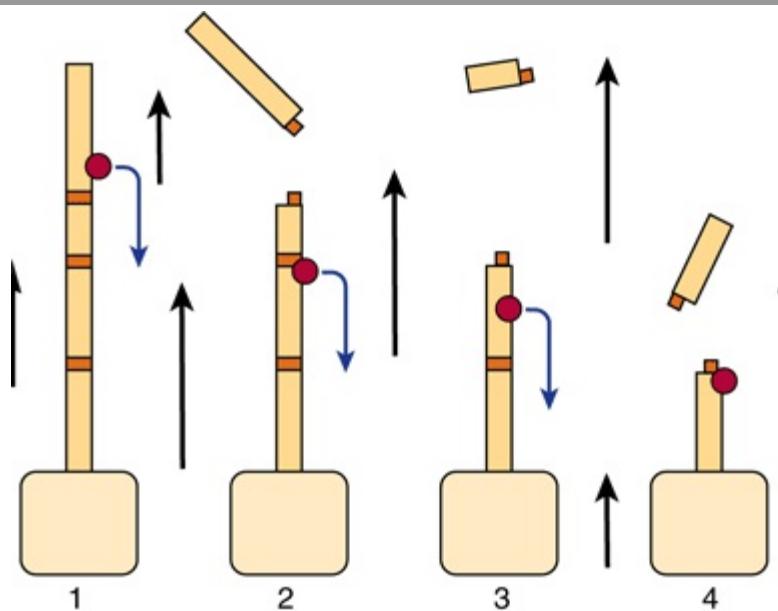
## THE CATALYTIC ACTIVITY OF ENZYMES FACILITATES THEIR DETECTION

The relatively small quantities of enzymes present in cells complicate determination of their presence and concentration. However, the amplification conferred by their ability to rapidly transform thousands of molecules of a specific substrate into products imbues each enzyme with the ability to reveal its presence. Assays of the catalytic activity of enzymes are frequently used in research and clinical laboratories. Under appropriate conditions (see Chapter 8), the rate of the catalytic reaction being monitored is proportionate to the amount of enzyme present, which allows its concentration to be inferred.

### Single-Molecule Enzymology

The limited sensitivity of traditional enzyme assays necessitates the use of a large group, or ensemble, of enzyme molecules in order to produce measurable quantities of product. The data obtained thus reflect the average catalytic capability of individual molecules. Recent advances in **nanotechnology** have made it possible to observe, usually by fluorescence microscopy, catalytic events involving individual enzyme and substrate molecules. Consequently, scientists can now measure the rate of single catalytic events and sometimes the individual steps in catalysis by a process called **single-molecule enzymology** (Figure 7–9).

**Figure 7–9**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Direct observation of single DNA cleavage events catalyzed by a restriction endonuclease.** DNA molecules immobilized to beads (pale yellow) are placed in a flowing stream of buffer (black arrows), which causes them to assume an extended conformation. Cleavage at one of the restriction sites (orange) by an endonuclease leads to a shortening of the DNA molecule, which can be observed directly in a microscope since the nucleotide bases in DNA are fluorescent. Although the endonuclease (red) does not fluoresce, and hence is invisible, the progressive manner in which the DNA molecule is shortened (1 → 4) reveals that the endonuclease binds to the free end of the DNA molecule and moves along it from site to site.

### Drug Discovery Requires Enzyme Assays Suitable for High-Throughput Screening

Enzymes constitute one of the primary classes of biomolecules targeted for the development of drugs and other therapeutic agents. Many antibiotics, for example, inhibit enzymes that are unique to microbial pathogens. The discovery of new drugs is greatly facilitated when a large number of potential pharmacophores can be assayed in a rapid, automated fashion—a process referred to as **high-throughput screening**. High-throughput screening (HTS) takes advantage of recent advances in robotics, optics, data processing, and microfluidics to conduct and analyze many thousands of simultaneous assays of the activity of a given enzyme. The most commonly used high-throughput screening devices employ 4–100  $\mu\text{L}$  volumes in 96, 384, or 1536 well plastic plates and fully automated equipment capable of dispensing substrates, coenzymes, enzymes, and potential inhibitors in a multiplicity of combinations and concentrations. High-throughput screening is ideal for surveying the numerous products of **combinatorial chemistry**, the simultaneous synthesis of large libraries of chemical compounds that contain all possible combinations of a set of chemical precursors. Enzyme assays that produce a chromogenic or fluorescent product are ideal, since optical detectors are readily engineered to permit the rapid analysis of multiple samples. At present, the sophisticated equipment required for truly large numbers of assays is available only in pharmaceutical houses, government-sponsored laboratories, and research universities. As described in Chapter 8, its principal use is the analysis of inhibitory compounds with ultimate potential for use as drugs.

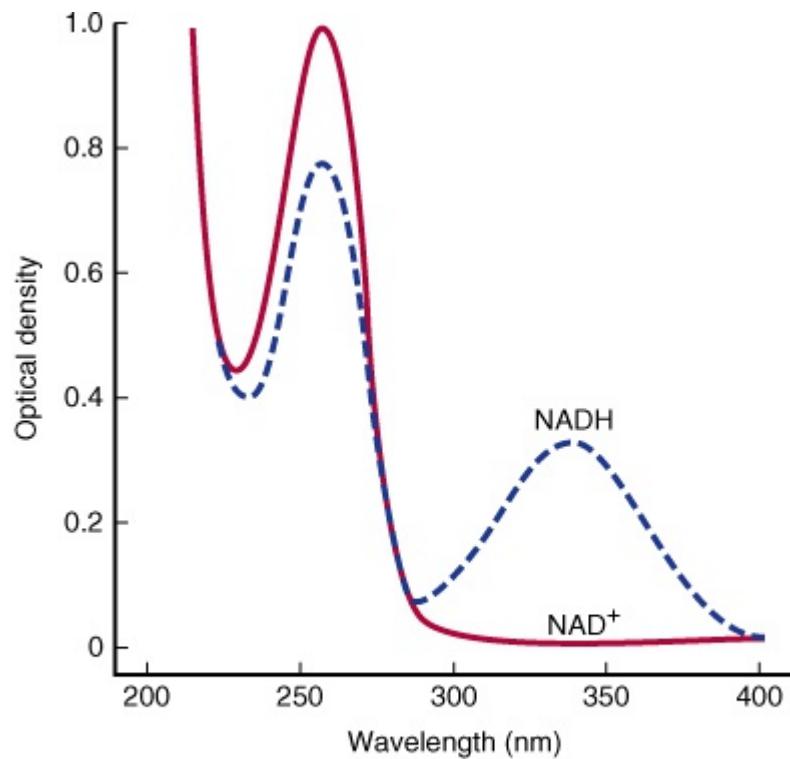
### Enzyme-Linked Immunoassays

The sensitivity of enzyme assays can be exploited to detect proteins that lack catalytic activity. **Enzyme-linked immunosorbent assays** (ELISAs) use antibodies covalently linked to a "reporter enzyme" such as alkaline phosphatase or horseradish peroxidase whose products are readily detected, generally by the absorbance of light or by fluorescence. Serum or other biologic samples to be tested are placed in a plastic microtiter plate, where the proteins adhere to the plastic surface and are immobilized. Any remaining absorbing areas of the well are then "blocked" by adding a nonantigenic protein such as bovine serum albumin. A solution of antibody covalently linked to a reporter enzyme is then added. The antibodies adhere to the immobilized antigen and are themselves immobilized. Excess free antibody molecules are then removed by washing. The presence and quantity of bound antibody is then determined by adding the substrate for the reporter enzyme.

### NAD(P)-Dependent Dehydrogenases Are Assayed Spectrophotometrically

The physicochemical properties of the reactants in an enzyme-catalyzed reaction dictate the options for the assay of enzyme activity. Spectrophotometric assays exploit the ability of a substrate or product to absorb light. The reduced coenzymes NADH and NADPH, written as NAD(P)H, absorb light at a wavelength of 340 nm, whereas their oxidized forms NAD(P)<sup>+</sup> do not (Figure 7–10). When NAD(P)<sup>+</sup> is reduced, the absorbance at 340 nm therefore increases in proportion to—and at a rate determined by—the quantity of NAD(P)H produced. Conversely, for a dehydrogenase that catalyzes the oxidation of NAD(P)H, a decrease in absorbance at 340 nm will be observed. In each case, the rate of change in optical density at 340 nm will be proportionate to the quantity of the enzyme present.

**Figure 7–10**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

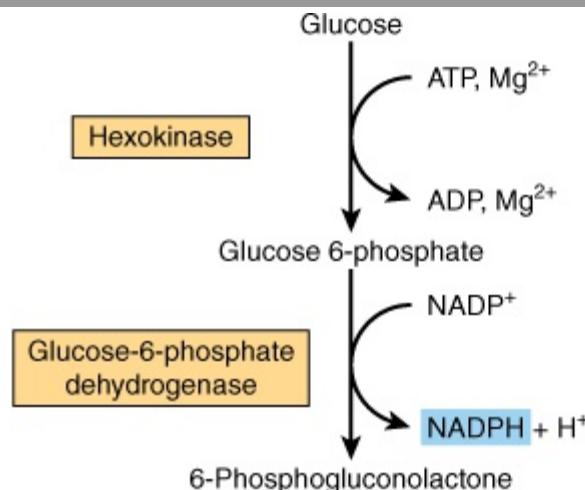
**Absorption spectra of NAD<sup>+</sup> and NADH.** Densities are for a 44 mg/L solution in a cell with a 1 cm light path. NADP<sup>+</sup> and NADPH have spectra analogous to NAD<sup>+</sup> and NADH, respectively.

### Many Enzymes Are Assayed by Coupling to a Dehydrogenase

The assay of enzymes whose reactions are not accompanied by a change in absorbance or fluorescence is generally more difficult. In some instances, the product or remaining substrate can be transformed into a more readily detected compound. In other instances, the reaction product may have to be separated from unreacted substrate prior to measurement. An alternative strategy is to devise a synthetic substrate whose product absorbs light or fluoresces. For example, *p*-nitrophenyl phosphate is an artificial substrate for certain phosphatases and for chymotrypsin that does not absorb visible light. However, following hydrolysis, the resulting *p*-nitrophenylate anion absorbs light at 419 nm.

Another quite general approach is to employ a "coupled" assay (Figure 7-11). Typically, a dehydrogenase whose substrate is the product of the enzyme of interest is added in catalytic excess. The rate of appearance or disappearance of NAD(P)H then depends on the rate of the enzyme reaction to which the dehydrogenase has been coupled.

**Figure 7-11**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Coupled enzyme assay for hexokinase activity.** The production of glucose 6-phosphate by hexokinase is coupled to the oxidation of this product by glucose-6-phosphate dehydrogenase in the presence of added enzyme and NADP<sup>+</sup>. When an excess of glucose-6-phosphate dehydrogenase is present, the rate of formation of NADPH, which can be measured at 340 nm, is governed by the rate of formation of glucose-6-phosphate by hexokinase.

## THE ANALYSIS OF CERTAIN ENZYMES AIDS DIAGNOSIS

The analysis of enzymes in blood plasma has played a central role in the diagnosis of several disease processes. Many enzymes are functional constituents of blood. Examples include pseudocholinesterase, lipoprotein lipase, and components of the cascade that trigger blood clotting and clot dissolution. Other enzymes are released into plasma following cell death or injury. While these latter enzymes perform no physiologic function in plasma, their appearance or levels can assist in the diagnosis and prognosis of diseases and injuries affecting specific tissues. Following injury, the plasma concentration of a released enzyme may rise early or late, and may decline rapidly or slowly. Proteins from the cytoplasm tend to appear more rapidly than those from subcellular organelles. The speed with which enzymes and other proteins are removed from plasma varies with their susceptibility to proteolysis and permeability through renal glomeruli.

Quantitative analysis of the activity of released enzymes or other proteins, typically in plasma or serum but also in urine or various cells, provides information concerning diagnosis, prognosis, and response to treatment. Assays of enzyme *activity* typically employ standard kinetic assays of initial reaction rates. **Table 7–2** lists several enzymes of value in clinical diagnosis. These enzymes are, however, not absolutely specific for the indicated disease. For example, elevated blood levels of prostatic acid phosphatase are associated typically with prostate cancer, but also with certain other cancers and noncancerous conditions. Consequently, enzyme assay data must be considered together with other factors elicited through a comprehensive clinical examination. Factors to be considered in interpreting enzyme data include patient age, sex, prior history, possible drug use, and the sensitivity and the diagnostic specificity of the enzyme test.

**Table 7–2 Principal Serum Enzymes Used in Clinical Diagnosis**

Serum Enzyme	Major Diagnostic Use
Aminotransferases	
Aspartate aminotransferase (AST, or SGOT)	Myocardial infarction
Alanine aminotransferase (ALT, or SGPT)	Viral hepatitis
Amylase	Acute pancreatitis
Ceruloplasmin	Hepatolenticular degeneration (Wilson's disease)
Creatine kinase	Muscle disorders and myocardial infarction
-Glutamyl transferase	Various liver diseases
Lactate dehydrogenase isozyme 5	Liver diseases
Lipase	Acute pancreatitis
Phosphatase, acid	Metastatic carcinoma of the prostate
Phosphatase, alkaline (isozymes)	Various bone disorders, obstructive liver diseases

**Note:** Many of the above enzymes are not specific to the disease listed.

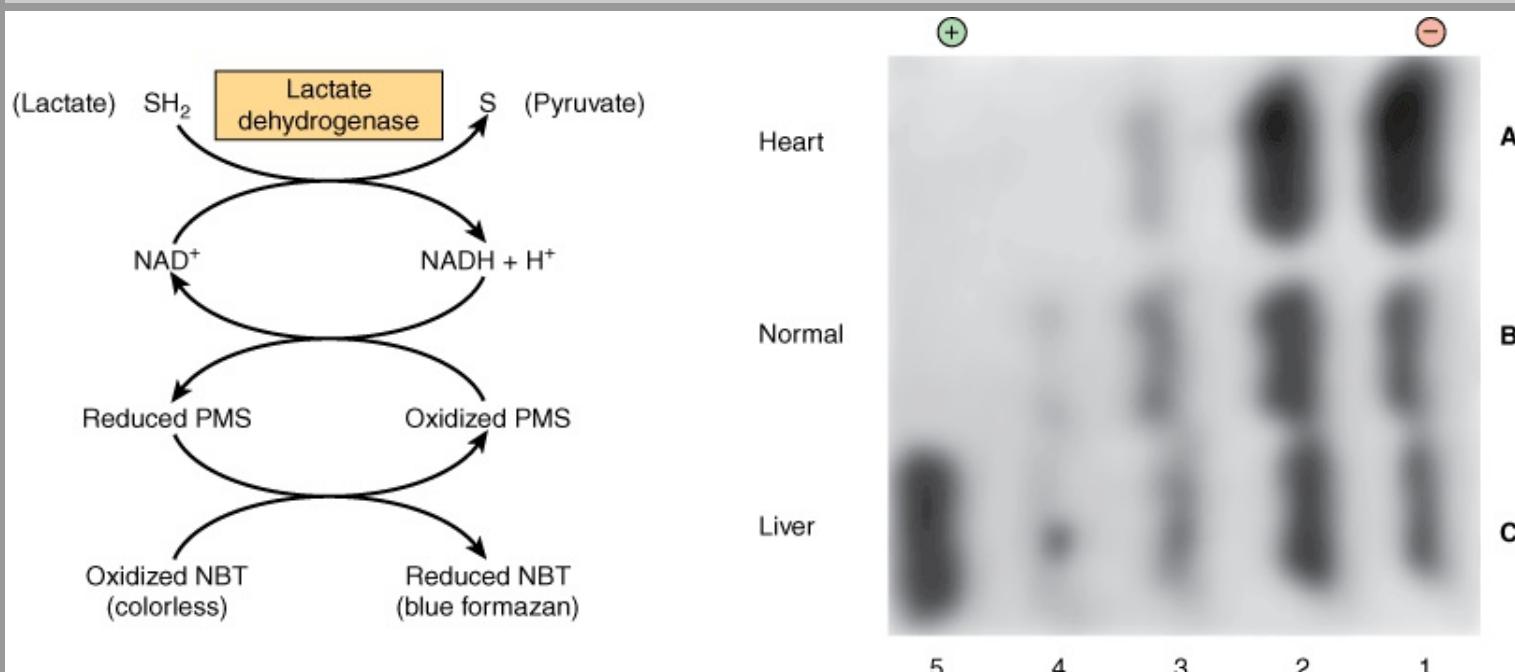
### Enzymes Assist Diagnosis of Myocardial Infarction

An enzyme useful for diagnostic enzymology should be relatively specific for the tissue or organ under study, should appear in the plasma or other fluid at a time useful for diagnosis (the "diagnostic window"), and should be amenable to automated assay. The enzymes used to confirm a myocardial infarction (MI) illustrate the concept of a "diagnostic window," and provide a historical perspective on the use of different enzymes for this purpose.

Detection of an enzyme must be possible within a few hours of an MI to confirm a preliminary diagnosis and permit initiation of appropriate therapy. Enzymes that only appear in the plasma 12 h or more following injury are thus of limited utility. The first enzymes used to diagnose MI were aspartate aminotransferase (AST), alanine aminotransferase (ALT), and lactate dehydrogenase. AST and ALT proved less than ideal, however, as they appear in plasma relatively slowly and are not specific to heart muscle. While LDH also is released relatively slowly into plasma, it offered the advantage of tissue specificity as a consequence of its quaternary structure.

Lactate dehydrogenase (LDH) is a tetrameric enzyme consisting of two monomer types: H (for heart) and M (for muscle) that combine to yield five LDH isozymes: HHHH ( $I_1$ ), HHHM ( $I_2$ ), HHMM ( $I_3$ ), HMMM ( $I_4$ ), and MMMM ( $I_5$ ). Tissue-specific expression of the H and M genes determines the relative proportions of each subunit in different tissues. Isozyme  $I_1$  predominates in heart tissue, and isozyme  $I_5$  in the liver. Thus, tissue injury releases a characteristic pattern of LDH isozymes that can be separated by electrophoresis and detected using a coupled assay (Figure 7–12). Today, LDH has been superseded as a marker for MI by other proteins that appear more rapidly in plasma.

**Figure 7–12**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Normal and pathologic patterns of lactate dehydrogenase (LDH) isozymes in human serum.** LDH isozymes of serum were separated by electrophoresis and visualized using the coupled reaction scheme shown on the left. (NBT, nitroblue tetrazolium; PMS, phenazine methylsulfate.) At right is shown the stained electropherogram. Pattern A is serum from a patient with a myocardial infarct; B is normal serum; and C is serum from a patient with liver disease. Arabic numerals denote specific LDH isozymes.

Creatine kinase (CK) has three isozymes: CK-MM (skeletal muscle), CK-BB (brain), and CK-MB (heart and skeletal muscle). CK-MB has a useful diagnostic window. It appears within 4–8 h of an MI, peaks at 24 h, and returns to baseline by 48–72 h. As for LDH, individual CK isozymes are separable by electrophoresis, thus facilitating detection. Assay of plasma CK levels continues in use to assess skeletal muscle disorders such as Duchene muscular dystrophy. Today, however, in most clinical laboratories the measurement of plasma troponin levels has

replaced CK as the preferred diagnostic marker for MI.

## Troponins

**Troponin** is a complex of three proteins involved in muscle contraction in *skeletal* and *cardiac muscle* but not in *smooth muscle* (see Chapter 49). Immunological measurement of plasma levels of cardiac troponins I and T provide sensitive and specific indicators of damage to heart muscle. Troponin levels rise for 2–6 h after an MI and remain elevated for 4–10 days. In addition to MI, other heart muscle damage also elevates serum troponin levels. Cardiac troponins thus serve as a marker of all heart muscle damage. The search for additional markers for heart disease, such as ischemia-modified albumin, and the simultaneous assessment of a spectrum of diagnostic markers via proteomics, continues to be an active area of clinical research.

Enzymes also can be employed in the clinical laboratory as tools for determining the concentration of critical metabolites. For example, glucose oxidase is frequently utilized to measure plasma glucose concentration. Enzymes are employed with increasing frequency as tools for the treatment of injury and disease. Tissue plasminogen activator (tPA) or streptokinase is used in the treatment of acute MI, while trypsin has been used in the treatment of cystic fibrosis (see Chapter 54).

## ENZYMES FACILITATE DIAGNOSIS OF GENETIC AND INFECTIOUS DISEASES

Many diagnostic techniques take advantage of the specificity and efficiency of the enzymes that act on oligonucleotides such as DNA (Chapter 39). Enzymes known as **restriction endonucleases**, for example, cleave double-stranded DNA at sites specified by a sequence of four, six, or more base pairs called **restriction sites**. Cleavage of a sample of DNA with a restriction enzyme produces a characteristic set of smaller DNA fragments (see Chapter 39). Deviations in the normal product pattern, called **restriction fragment length polymorphisms (RFLPs)**, occur if a mutation renders a restriction site unrecognizable to its cognate restriction endonuclease or, alternatively, generates a new recognition site. RFLPs are currently utilized to facilitate prenatal detection of a number of hereditary disorders, including sickle cell trait, beta-thalassemia, infant phenylketonuria, and Huntington's disease.

The **polymerase chain reaction (PCR)** employs a thermostable DNA polymerase and appropriate oligonucleotide primers to produce thousands of copies of a defined segment of DNA from a minute quantity of starting material (see Chapter 39). PCR enables medical, biological, and forensic scientists to detect and characterize DNA present initially at levels too low for direct detection. In addition to screening for genetic mutations, PCR can be used to detect and identify pathogens and parasites such as *Trypanosoma cruzi*, the causative agent of Chagas' disease, and *Neisseria meningitidis*, the causative agent of bacterial meningitis, through the selective amplification of their DNA.

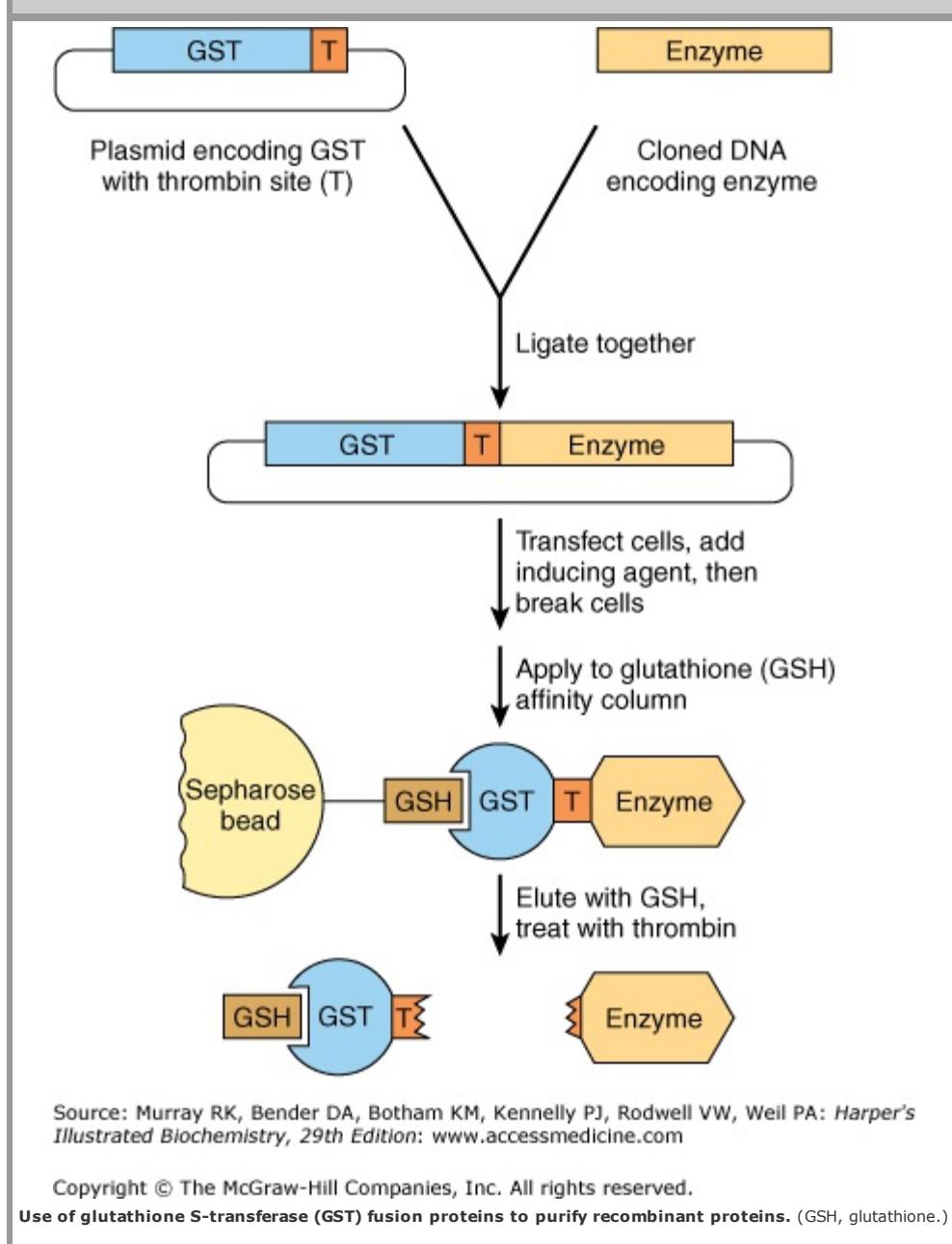
## RECOMBINANT DNA PROVIDES AN IMPORTANT TOOL FOR STUDYING ENZYMES

Recombinant DNA technology has emerged as an important asset in the study of enzymes. Highly purified samples of enzymes are necessary for the study of their structure and function. The isolation of an individual enzyme, particularly one present in low concentration, from among the thousands of proteins present in a cell can be extremely difficult. If the gene for the enzyme of interest has been cloned, it generally is possible to produce large quantities of its encoded protein in *Escherichia coli* or yeast. However, not all animal proteins can be expressed in an active form in microbial cells, nor do microbes perform certain post-translational processing tasks. For these reasons, a gene may be expressed in cultured animal cell systems employing the baculovirus expression vector to transform cultured insect cells. For more details concerning recombinant DNA techniques, see Chapter 39.

### Recombinant Fusion Proteins Are Purified by Affinity Chromatography

Recombinant DNA technology can also be used to create modified proteins that are readily purified by affinity chromatography. The gene of interest is linked to an oligonucleotide sequence that encodes a carboxyl or amino terminal extension to the encoded protein. The resulting modified protein, termed as a **fusion protein**, contains a domain tailored to interact with a specific affinity support. One popular approach is to attach an oligonucleotide that encodes six consecutive histidine residues. The expressed "His tag" protein binds to chromatographic supports that contain an immobilized divalent metal ion such as  $\text{Ni}^{2+}$  or  $\text{Cd}^{2+}$ . Alternatively, the substrate-binding domain of glutathione S-transferase (GST) can serve as a "GST tag." **Figure 7–13** illustrates the purification of a GST-fusion protein using an affinity support containing bound glutathione. Fusion proteins also often encode a cleavage site for a highly specific protease such as thrombin in the region that links the two portions of the protein. This permits removal of the added fusion domain following affinity purification.

**Figure 7–13**



### Site-Directed Mutagenesis Provides Mechanistic Insights

Once the ability to express a protein from its cloned gene has been established, it is possible to employ **site-directed mutagenesis** to change specific aminoacyl residues by altering their codons. Used in combination with kinetic analyses and x-ray crystallography, this approach facilitates identification of the specific roles of given aminoacyl residues in substrate binding and catalysis. For example, the inference that a particular aminoacyl residue functions as a general acid can be tested by replacing it with an aminoacyl residue incapable of donating a proton.

## RIBOZYMES: ARTIFACTS FROM THE RNA WORLD

### Cech Discovered the First Catalytic RNA Molecule

The participation of enzyme catalysts in the posttranslational maturation of certain proteins has analogies in the RNA world. Many RNA molecules undergo processing that both removes segments of oligonucleotide and re-ligates the remaining segments to form the mature product (Chapter 36). Not all of these catalysts are proteins, however. While examining the processing of ribosomal RNA (rRNA) molecules in the ciliated protozoan *Tetrahymena*, Thomas Cech and his coworkers observed, in the early 1980s, that processing of the 26S rRNA proceeded smoothly *in vitro* even in the total *absence* of protein. The source of this splicing activity was traced to a 413 bp catalytic segment that retained its catalytic activity even when replicated in *E. coli* (Chapter 39). Prior to that time, polynucleotides had been thought to serve solely as information storage and transmission entities, and that catalysis was restricted solely to proteins.

Several other ribozymes have since been discovered. The vast majority catalyze nucleophilic displacement reactions that target the phosphodiester bonds of the RNA backbone. In small self-cleaving RNAs, such as hammerhead or hepatitis delta virus RNA, the attacking nucleophile is water and the result is hydrolysis. For the large group I intron ribozymes, the attacking nucleophile is the 3'-hydroxyl of the terminal ribose of another segment of RNA and the result is a splicing reaction.

### The Ribosome—the Ultimate Ribozyme

The ribozyme was the first recognized "molecular machine." A massive complex comprised of scores of protein subunits and several large ribosomal RNA molecules, the ribosome performs the vitally important and highly complex process of synthesizing long polypeptide chains following the instructions encoded in messenger RNA molecules (Chapter 37). For many years, it was assumed that ribosomal RNAs played a passive, structural role, or perhaps helped in the recognition of cognate mRNAs through a base pairing mechanism.

### The RNA World Hypothesis

The discovery of ribozymes had a profound influence on evolutionary theory. For many years, scientists had hypothesized that the first biologic catalysts were formed when amino acids contained in the primordial soup coalesced to form the first simple proteins. With the realization that RNA could both carry information and catalyze simple chemical reactions, a new "RNA World" hypothesis emerged in which RNA constituted the first biological macromolecule. Eventually, DNA emerged as a more chemically stable oligonucleotide for long-term information storage while proteins, by virtue of their much greater variety of chemical functional groups, dominated catalysis. If one assumes that some sort of RNA-protein hybrid was formed as an intermediate in the transition from ribonucleotide to polypeptide catalysts, one need look no further than the ribosome to find the presumed missing link.

Why did not proteins take over all catalytic functions? Presumably, in the case of the ribosome the process was both too complex and too essential to permit much opportunity for possible competitors to gain a foothold. In the case of the small self-cleaving RNAs and self-splicing introns, they may represent one of the few cases in which RNA autocatalysis is more efficient than development of a new protein catalyst.

## SUMMARY

- Enzymes are efficient catalysts whose stringent specificity extends to the kind of reaction catalyzed, and typically to a single substrate.
- Organic and inorganic prosthetic groups, cofactors, and coenzymes play important roles in catalysis. Coenzymes, many of which are derivatives of B vitamins, serve as "shuttles" for commonly-used groups such as amines, electrons, and acetyl groups.
- During catalysis, enzymes frequently redirect the conformational changes induced by substrate binding to effect complementary changes in the substrate that facilitate its transformation into product.
- Catalytic mechanisms employed by enzymes include the introduction of strain, approximation of reactants, acid-base catalysis, and covalent catalysis. HIV protease illustrates acid-base catalysis; chymotrypsin and fructose-2,6-bisphosphatase illustrate covalent catalysis.
- Aminoacyl residues that participate in catalysis are highly conserved among all classes of a given enzyme. Site-directed mutagenesis, used to change residues suspected of being important in catalysis or substrate binding, provides insights into mechanisms of enzyme action.
- The catalytic activity of enzymes reveals their presence, facilitates their detection, and provides the basis for enzyme-linked immunoassays. Many enzymes can be assayed spectrophotometrically by coupling them to an NAD(P)<sup>+</sup>-dependent dehydrogenase.
- Combinatorial chemistry generates extensive libraries of potential enzyme activators and inhibitors that can be tested by high-throughput screening.
- Assay of plasma enzymes aids diagnosis and prognosis of myocardial infarction, acute pancreatitis, and various bone and liver disorders.
- Restriction endonucleases facilitate diagnosis of genetic diseases by revealing restriction fragment length polymorphisms, and the polymerase chain reaction (PCR) amplifies DNA initially present in quantities too small for analysis.
- Attachment of a polyhistidyl, glutathione S-transferase (GST), or other "tag" to the N- or C-terminus of a recombinant protein facilitates its purification by affinity chromatography on a solid support that contains an immobilized ligand such as a divalent cation (eg, Ni<sup>2+</sup>) or GST. Specific proteases can then remove affinity "tags" and generate the native enzyme.
- Not all enzymes are proteins. Several ribozymes are known that can cut and re-splice the phosphodiester bonds of RNA. In the ribosome, it is the rRNA and not the polypeptide components that are primarily responsible for catalysis.

## REFERENCES

- Brik A, Wong C-H: HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem* 2003;1:5.
- Burtis CA, Ashwood ER, Bruns DE: *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 4th ed. Elsevier, 2006.
- Cornish PV, Ha T: A survey of single-molecule techniques in chemical biology. *ACS Chem Biol* 2007;2:53.
- Doudna JA, Lorsch JR: Ribozyme catalysis: not different, just worse. *Nature Struct Biol* 2005;12:395.
- Frey PA, Hegeman AD: *Enzyme Reaction Mechanisms*. Oxford University Press. 2006.
- Geyser HM, Schoenen F, Wagner D, Wagner R: Combinatorial compound libraries for drug discovery: an ongoing challenge. *Nature Rev Drug Disc* 2003;2:222.
- Goddard J-P, Reymond J-L: Enzyme assays for high-throughput screening. *Curr Opin Biotech* 2004;15:314.
- Gupta S, de Lemos JA: Use and misuse of cardiac troponins in clinical practice. *Prog Cardiovasc Dis* 2007;50:151.
- Hedstrom L: Serine protease mechanism and specificity. *Chem Rev* 2002;102:4501.
- Melanson SF, Tanasijevic MJ: Laboratory diagnosis of acute myocardial injury. *Cardiovascular Pathol* 2005;14:156.
- Pereira DA, Williams JA: Origin and evolution of high throughput screening. *Br J Pharmacol* 2007;152:53.
- René AWF, Titman CM, Pratap CV, et al: A molecular switch and proton wire synchronize the active sites in thiamine enzymes. *Science* 2004;306:872.
- Schmeing TM, Ramakrishnan V: What recent ribosome structures have revealed about the mechanism of translation. *Nature* 2009; 461:1234.
- Schafer B, Gemeinhardt H, Greulich KO: Direct microscopic observation of the time course of single-molecule DNA restriction reactions. *Agnew Chem Int Ed* 2001;40:4663.
- Silverman RB: *The Organic Chemistry of Enzyme-Catalyzed Reactions*. Academic Press, 2002.
- Sundaresan V, Abrol R: Towards a general model for protein-substrate stereoselectivity. *Protein Sci* 2002;11:1330.
- Todd AE, Orengo CA, Thornton JM: Plasticity of enzyme active sites. *Trends Biochem Sci* 2002;27:419.
- Urich T, Gomes CM, Kletzin A, et al: X-ray structure of a self-compartmentalizing sulfur cycle metalloenzyme. *Science* 2006;311:996.
- Walsh CT: *Enzymatic Reaction Mechanisms*. Freeman, 1979.

## OBJECTIVES

After studying this chapter, you should be able to:

- Describe the scope and overall purposes of the study of enzyme kinetics.
- Indicate whether  $\Delta G$ , the overall change in free energy for a reaction, is dependent on reaction mechanism.
- Indicate whether  $\Delta G$  is a function of the rates of reactions.
- Explain the relationship between  $K_{eq}$ , concentrations of substrates and products at equilibrium, and the ratio of the rate constants  $k_1/k_{-1}$ .
- Outline how temperature and the concentration of hydrogen ions, enzyme, and substrate affect the rate of an enzyme-catalyzed reaction.
- Indicate why laboratory measurement of the rate of an enzyme-catalyzed reaction typically employs initial rate conditions.
- Describe the application of linear forms of the Michaelis-Menten equation to the determination of  $K_m$  and  $V_{max}$ .
- Give one reason why a linear form of the Hill equation is used to evaluate the substrate-binding kinetics exhibited by some multimeric enzymes.
- Contrast the effects of an increasing concentration of substrate on the kinetics of simple competitive and noncompetitive inhibition.
- Describe the ways in which substrates add to, and products depart from, an enzyme that follows a ping-pong mechanism and do the same for an enzyme that follows a rapid-equilibrium mechanism.
- Illustrate the utility of enzyme kinetics in ascertaining the mode of action of drugs.

## BIOMEDICAL IMPORTANCE

**Enzyme kinetics** is the field of biochemistry concerned with the quantitative measurement of the rates of enzyme-catalyzed reactions and the systematic study of factors that affect these rates. Kinetic analysis can reveal the number and order of the individual steps by which enzymes transform substrates into products. Together with site-directed mutagenesis and other techniques that probe the protein structure, kinetic analyses can reveal details of the catalytic mechanism of a given enzyme.

A complete, balanced set of enzyme activities is of fundamental importance for maintaining homeostasis. An understanding of enzyme kinetics thus is important to understanding how physiologic stresses such as anoxia, metabolic acidosis or alkalosis, toxins, and pharmacologic agents affect that balance. The involvement of enzymes in virtually all physiologic processes makes them the targets of choice for drugs that cure or ameliorate human disease. Applied enzyme kinetics represents the principal tool by which scientists identify and characterize therapeutic agents that selectively inhibit the rates of specific enzyme-catalyzed processes. Enzyme kinetics thus plays a central and critical role in drug discovery and comparative pharmacodynamics, as well as in elucidating the mode of action of drugs.

## CHEMICAL REACTIONS ARE DESCRIBED USING BALANCED EQUATIONS

A **balanced chemical equation** lists the initial chemical species (substrates) present and the new chemical species (products) formed for a particular chemical reaction, all in their correct proportions or **stoichiometry**. For example, balanced equation (1) describes the reaction of one molecule each of substrates A and B to form one molecule each of products P and Q:



The double arrows indicate reversibility, an intrinsic property of all chemical reactions. Thus, for reaction (1), if A and B can form P and Q, then P and Q can also form A and B. Designation of a particular reactant as a "substrate" or "product" is therefore somewhat arbitrary since the products for a reaction written in one direction are the substrates for the reverse reaction. The term "products" is, however, often used to designate the reactants whose formation is thermodynamically favored. Reactions for which thermodynamic factors strongly favor formation of the products to which the arrow points often are represented with a single arrow as if they were "irreversible":



Unidirectional arrows are also used to describe reactions in living cells where the products of reaction (2) are immediately consumed by a subsequent enzyme-catalyzed reaction. The rapid removal of product P or Q therefore effectively precludes occurrence of the reverse reaction, rendering equation (2) **functionally irreversible under physiologic conditions**.

## CHANGES IN FREE ENERGY DETERMINE THE DIRECTION & EQUILIBRIUM STATE OF CHEMICAL REACTIONS

The Gibbs free energy change  $\Delta G$  (also called either free energy or Gibbs energy) describes both the *direction* in which a chemical reaction will tend to proceed and the concentrations of reactants and products that will be present at equilibrium.  $\Delta G$  for a chemical reaction equals the sum of the free energies of formation of the reaction products  $\Delta G_p$  minus the sum of the free energies of formation of the substrates  $\Delta G_s$ .  $\Delta G^0$  denotes the change in free energy that accompanies transition from the standard state, one-molar concentrations of substrates and products, to equilibrium. A more useful biochemical term is  $\Delta G'$ , which defines  $\Delta G^0$  at a standard state of  $10^{-7}$  M protons, pH 7.0 (Chapter 11). If the free energy of formation of the products is lower than that of the substrates, the signs of  $\Delta G^0$  and  $\Delta G'$  will be negative, indicating that the reaction as written is favored in the direction left to right. Such reactions are referred to as **spontaneous**. The **sign** and the **magnitude** of the free energy change determine how far the reaction will proceed. Equation (3) illustrates the relationship between the equilibrium constant  $K_{eq}$  and  $\Delta G^0$ :

$$\Delta G^0 = -RT \ln K_{eq} \quad (3)$$

where  $R$  is the gas constant (1.98 cal/mol°K or 8.31 J/mol°K) and  $T$  is the absolute temperature in degrees Kelvin.  $K_{eq}$  is equal to the product of the concentrations of the reaction products, each raised to the power of their stoichiometry, divided by the product of the substrates, each raised to the power of their stoichiometry:



$$K_{eq} = \frac{[P][Q]}{[A][B]} \quad (4)$$

and for reaction (5)



$$K_{eq} = \frac{[P]}{[A]^2} \quad (6)$$

$\Delta G^0$  may be calculated from equation (3) if the molar concentrations of substrates and products present at equilibrium are known. If  $\Delta G^0$  is a negative number,  $K_{eq}$  will be greater than unity, and the concentration of products at equilibrium will exceed that of the substrates. If  $\Delta G^0$  is positive,  $K_{eq}$  will be less than unity, and the formation of substrates will be favored.

Note that, since  $\Delta G^0$  is a function exclusively of the initial and final states of the reacting species, it can provide information only about the *direction* and *equilibrium state* of the reaction.  $\Delta G^0$  is independent of the **mechanism** of the reaction and therefore provides no information concerning **rates** of reactions. Consequently—and as explained below—although a reaction may have a large negative  $\Delta G^0$  or  $\Delta G'$ , it may nevertheless take place at a negligible rate.

## THE RATES OF REACTIONS ARE DETERMINED BY THEIR ACTIVATION ENERGY

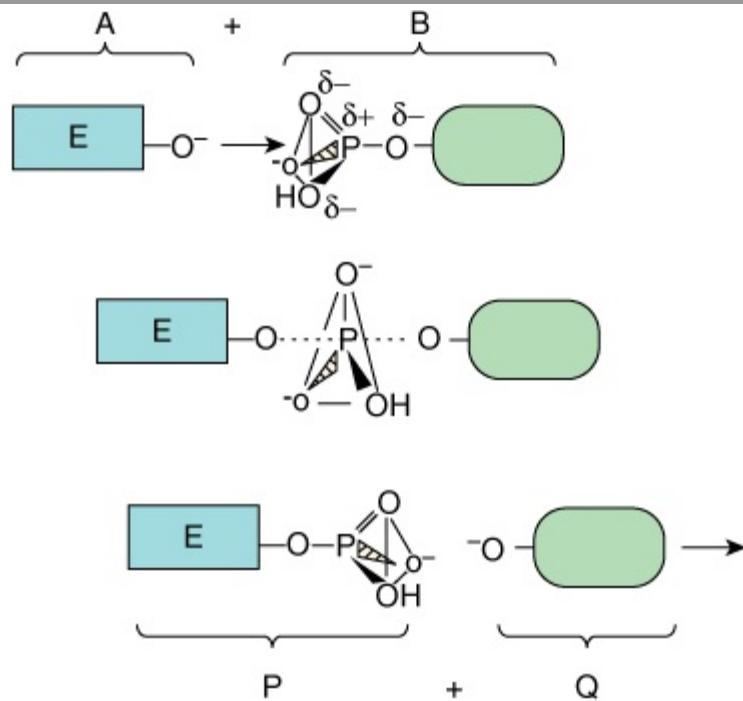
### Reactions Proceed Via Transition States

The concept of the **transition state** is fundamental to understanding the chemical and thermodynamic basis of catalysis. Equation (7) depicts a group transfer reaction in which an entering group E displaces a leaving group L, attached initially to R:



The net result of this process is to transfer group R from L to E. Midway through the displacement, the bond between R and L has weakened but has not yet been completely severed, and the new bond between E and R is yet incompletely formed. This transient intermediate—in which neither free substrate nor product exists—is termed the **transition state**,  $E \cdots R \cdots L$ . Dotted lines represent the "partial" bonds that are undergoing formation and rupture. **Figure 8-1** provides a more detailed illustration of the transition state intermediate formed during the transfer of a phosphoryl group.

**Figure 8-1**

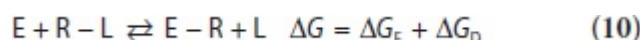


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Formation of a transition state intermediate during a simple chemical reaction,  $A + B \rightleftharpoons P + Q$ .** Shown are three stages of a chemical reaction in which a phosphoryl group is transferred from leaving group L to entering group E (A). Top: entering group E (A) approaches the other reactant, L-phosphate (B). Notice how the three oxygen atoms linked by the triangular lines and the phosphorus atom of the phosphoryl group form a pyramid. Center: as E approaches L-phosphate, the new bond between E and the phosphate group begins to form (dotted line) as that linking L to the phosphate group weakens. These partially formed bonds are indicated by dotted lines. Bottom: formation of the new product, E-phosphate (P), is now complete as the leaving group L (Q) exits. Notice how the geometry of the phosphoryl group differs between the transition state and the substrate or product. Notice how the phosphorus and three oxygen atoms that occupy the four corners of a pyramid in the substrate and product become coplanar, as emphasized by the triangle, in the transition state.

Reaction (7) can be thought of as consisting of two "partial reactions," the first corresponding to the formation (F) and the second to the subsequent decay (D) of the transition state intermediate. As for all reactions, characteristic changes in free energy,  $\Delta G_F$  and  $\Delta G_D$  are associated with each partial reaction:



For the overall reaction (10),  $\Delta G$  is the sum of  $\Delta G_F$  and  $\Delta G_D$ . As for any equation of two terms, it is not possible to infer from  $\Delta G$  either the sign or the magnitude of  $\Delta G_F$  or  $\Delta G_D$ .

Many reactions involve multiple transition states, each with an associated change in free energy. For these reactions, the overall  $\Delta G$  represents the sum of *all* of the free energy changes associated with the formation and decay of *all* of the transition states. **Therefore, it is not possible to infer from the overall  $\Delta G$  the number or type of transition states through which the reaction proceeds.** Stated another way, overall thermodynamics tells us nothing about kinetics.

### $\Delta G_F$ Defines the Activation Energy

Regardless of the sign or magnitude of  $\Delta G$ ,  $\Delta G_F$  for the overwhelming majority of chemical reactions has a positive sign. The formation of transition state intermediates therefore requires surmounting energy barriers. For this reason,  $\Delta G_F$  for reaching a transition state is often termed the **activation energy**,  $E_{act}$ . The ease—and hence the frequency—with which this barrier is overcome is inversely related to  $E_{act}$ . The thermodynamic parameters that determine how *fast* a reaction proceeds thus are the  $\Delta G_F$  values for formation of the transition states through which the reaction proceeds. For a simple reaction, where  $\propto$  means "proportionate to,"

$$\text{Rate} \propto e^{-E_{act}/RT} \quad (11)$$

The activation energy for the reaction proceeding in the opposite direction to that drawn is equal to  $-\Delta G_D$ .

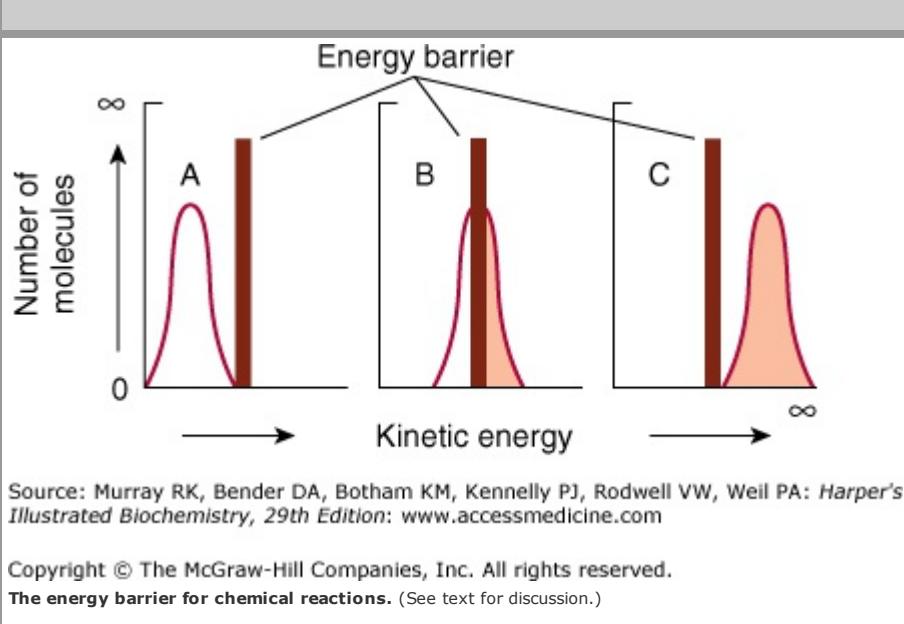
## NUMEROUS FACTORS AFFECT THE REACTION RATE

The **kinetic theory**—also called the **collision theory**—of chemical kinetics states that for two molecules to react they (1) must approach within bond-forming distance of one another, or "collide," and (2) must possess sufficient kinetic energy to overcome the energy barrier for reaching the transition state. It therefore follows that anything that increases the frequency or energy of collision between substrates will increase the rate of the reaction in which they participate.

### Temperature

Raising the temperature increases the kinetic energy of molecules. As illustrated in **Figure 8–2**, the total number of molecules whose kinetic energy exceeds the energy barrier  $E_{act}$  (vertical bar) for formation of products increases from low (A) through intermediate (B) to high (C) temperatures. Increasing the kinetic energy of molecules also increases their rapidity of motion and therefore the frequency with which they collide. This combination of more frequent and more highly energetic, and hence productive, collisions increases the reaction rate.

**Figure 8–2**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**The energy barrier for chemical reactions.** (See text for discussion.)

### Reactant Concentration

The frequency with which molecules collide is directly proportionate to their concentrations. For two different molecules A and B, the frequency with which they collide will double if the concentration of either A or B is doubled. If the concentrations of both A and B are doubled, the probability of collision will increase fourfold.

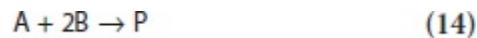
For a chemical reaction proceeding at constant temperature that involves one molecule each of A and B,



the number of the molecules that possess kinetic energy sufficient to overcome the activation energy barrier will be a constant. The number of collisions with sufficient energy to produce product P therefore will be directly proportionate to the number of collisions between A and B, and thus to their molar concentrations, denoted by the square brackets:

$$\text{Rate} \propto [A][B] \quad (13)$$

Similarly, for the reaction represented by



which can also be written as



The corresponding rate expression is

$$\text{Rate} \propto [A][B][B] \quad (16)$$

or

$$\text{Rate} \propto [A][B]^2 \quad (17)$$

For the general case, when  $n$  molecules of A react with  $m$  molecules of B,



the rate expression is

$$\text{Rate} \propto [A]^n[B]^m \quad (19)$$

Replacing the proportionality sign with an equals sign by introducing a **rate constant  $k$**  characteristic of the reaction under study gives equations (20) and (21), in which the subscripts 1 and  $-1$  refer to the forward and reverse reactions, respectively:

$$\text{Rate}_1 = k_1[A]^n[B]^m \quad (20)$$

$$\text{Rate}_{-1} = k_{-1}[P] \quad (21)$$

The sum of the molar ratios of the reactants defines the **kinetic order** of the reaction. Consider reaction (5). The stoichiometric coefficient for the sole reactant, A, is 2. Therefore, the rate of production of P is proportional to the square of [A] and the reaction is said to be *second order* with respect to reactant A. In this instance, the overall reaction is also *second order*. Therefore,  $k_1$  is referred to as a *second-order rate constant*.

Reaction (12) describes a simple second-order reaction between two different reactants, A and B. The stoichiometric coefficient for each reactant is 1. Therefore, while the overall order of the reaction is 2, it is said to be *first order* with respect to A and *first order* with respect to B. In the laboratory, the kinetic order of a reaction with respect to a particular reactant, referred to as the variable reactant or substrate, can be determined by maintaining the concentration of the other reactants at a constant, or fixed, concentration in large excess over the variable reactant. Under these *pseudo-first-order conditions*, the concentration of the fixed reactant(s) remains virtually constant. Thus, the rate of reaction will depend exclusively on the concentration of the variable reactant, sometimes also called the limiting reactant. The concepts of reaction order and pseudo-first-order conditions apply not only to simple chemical reactions but also to enzyme-catalyzed reactions.

### $K_{eq}$ Is a Ratio of Rate Constants

While all chemical reactions are to some extent reversible, at equilibrium the *overall* concentrations of reactants and products remain constant. At equilibrium, the rate of conversion of substrates to products therefore equals the rate at which products are converted to substrates:

$$\text{Rate}_1 = \text{Rate}_{-1} \quad (22)$$

Therefore,

$$k_1[A]^r[B]^m = k_{-1}[P] \quad (23)$$

and

$$\frac{k_1}{k_{-1}} = \frac{[P]}{[A]^r[B]^m} \quad (24)$$

The ratio of  $k_1$  to  $k_{-1}$  is termed the equilibrium constant,  $K_{\text{eq}}$ . The following important properties of a system at equilibrium must be kept in mind.

1. The equilibrium constant is a ratio of the reaction *rate constants* (not the reaction *rates*).
2. At equilibrium, the reaction *rates* (not the *rate constants*) of the forward and back reactions are equal.
3. Equilibrium is a *dynamic* state. Although there is no *net* change in the concentration of substrates or products, individual substrate and product molecules are continually being interconverted.
4. The numeric value of the equilibrium constant  $K_{\text{eq}}$  can be calculated either from the concentrations of substrates and products at equilibrium or from the ratio  $k_1/k_{-1}$ .

## THE KINETICS OF ENZYMATIC CATALYSIS

### Enzymes Lower the Activation Energy Barrier for a Reaction

All enzymes accelerate reaction rates by lowering  $\Delta G_F$  for the formation of transition states. However, they may differ in the way this is achieved. Where the mechanism or the sequence of chemical steps at the active site is essentially equivalent to those for the same reaction proceeding in the absence of a catalyst, **the environment of the active site lowers  $\Delta G_F$**  by stabilizing the transition state intermediates. To put it another way, the enzyme can be envisioned as binding to the transition state intermediate (**Figure 8–1**) more tightly than it does to either substrates or products. As discussed in Chapter 7, stabilization can involve (1) acid–base groups suitably positioned to transfer protons to or from the developing transition state intermediate, (2) suitably positioned charged groups or metal ions that stabilize developing charges, or (3) the imposition of steric strain on substrates so that their geometry approaches that of the transition state. HIV protease (see **Figure 7–6**) illustrates catalysis by an enzyme that lowers the activation barrier by stabilizing a transition state intermediate.

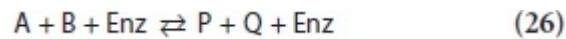
Catalysis by enzymes that proceeds via a *unique* reaction mechanism typically occurs when the transition state intermediate forms a covalent bond with the enzyme (**covalent catalysis**). The catalytic mechanism of the serine protease chymotrypsin (see **Figure 7–7**) illustrates how an enzyme utilizes covalent catalysis to provide a unique reaction pathway.

## ENZYMES DO NOT AFFECT $K_{\text{eq}}$

While enzymes undergo transient modifications during the process of catalysis, they always emerge unchanged at the completion of the reaction. **The presence of an enzyme therefore has no effect on  $\Delta G^{\circ}$  for the overall reaction**, which is a function solely of the **initial and final states** of the reactants. Equation (25) shows the relationship between the equilibrium constant for a reaction and the standard free energy change for that reaction:

$$\Delta G^{\circ} = -RT \ln K_{\text{eq}} \quad (25)$$

This principle is perhaps most readily illustrated by including the presence of the enzyme (Enz) in the calculation of the equilibrium constant for an enzyme-catalyzed reaction:



Since the enzyme on both sides of the double arrows is present in equal quantity and identical form, the expression for the equilibrium constant,

$$K_{\text{eq}} = \frac{[P][Q][\text{Enz}]}{[A][B][\text{Enz}]} \quad (27)$$

reduces to one identical to that for the reaction in the absence of the enzyme:

$$K_{\text{eq}} = \frac{[P][Q]}{[A][B]} \quad (28)$$

Enzymes therefore have no effect on  $K_{\text{eq}}$ .

## MULTIPLE FACTORS AFFECT THE RATES OF ENZYME-CATALYZED REACTIONS

### Temperature

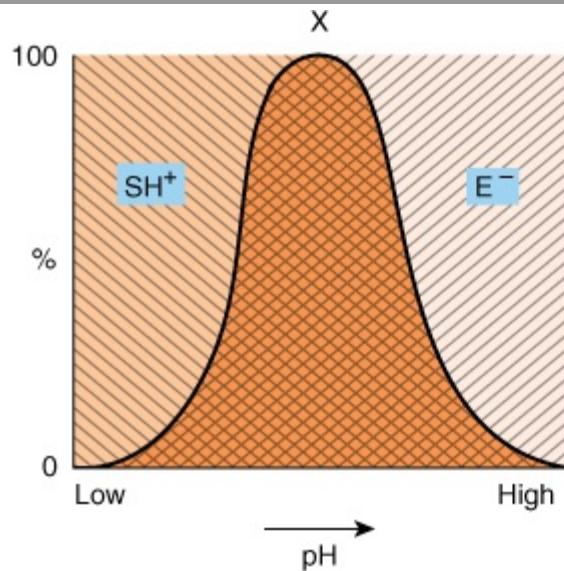
Raising the temperature increases the rate of both uncatalyzed and enzyme-catalyzed reactions by increasing the kinetic energy and the collision frequency of the reacting molecules. However, heat energy can also increase the kinetic energy of the enzyme to a point that exceeds the energy barrier for disrupting the noncovalent interactions that maintain its three-dimensional structure. The polypeptide chain then begins to unfold, or **denature**, with an accompanying loss of the catalytic activity. The temperature range over which an enzyme maintains a stable, catalytically competent conformation depends upon—and typically moderately exceeds—the normal temperature of the cells in which it resides. Enzymes from humans generally exhibit stability at temperatures up to 45–55°C. By contrast, enzymes from the thermophilic microorganisms that reside in volcanic hot springs or undersea hydrothermal vents may be stable at temperatures up to or even above 100°C.

The **temperature coefficient (Q<sub>10</sub>)** is the factor by which the rate of a biologic process increases for a 10°C increase in temperature. For the temperatures over which enzymes are stable, the rates of most biological processes typically double for a 10°C rise in temperature ( $Q_{10} = 2$ ). Changes in the rates of enzyme-catalyzed reactions that accompany a rise or fall in body temperature constitute a prominent survival feature for "cold-blooded" life forms such as lizards or fish, whose body temperatures are dictated by the external environment. However, for mammals and other homeothermic organisms, changes in enzyme reaction rates with temperature assume physiologic importance only in circumstances such as fever or hypothermia.

### Hydrogen Ion Concentration

The rate of almost all enzyme-catalyzed reactions exhibits a significant dependence on hydrogen ion concentration. Most intracellular enzymes exhibit optimal activity at pH values between 5 and 9. The relationship of activity to hydrogen ion concentration (Figure 8–3) reflects the balance between enzyme denaturation at high or low pH and effects on the charged state of the enzyme, the substrates, or both. For enzymes whose mechanism involves acid-base catalysis, the residues involved must be in the appropriate state of protonation for the reaction to proceed. The binding and recognition of substrate molecules with dissociable groups also typically involves the formation of salt bridges with the enzyme. The most common charged groups are carboxylate groups (negative) and protonated amines (positive). Gain or loss of critical charged groups adversely affects substrate binding and thus will retard or abolish catalysis.

**Figure 8–3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Effect of pH on enzyme activity.** Consider, for example, a negatively charged enzyme ( $E^-$ ) that binds a positively charged substrate ( $SH^+$ ). Shown is the proportion (%) of  $SH^+$  [\\] and of  $E^-$  [/] as a function of pH. Only in the cross-hatched area do both the enzyme and the substrate bear an appropriate charge.

## ASSAYS OF ENZYME-CATALYZED REACTIONS TYPICALLY MEASURE THE INITIAL VELOCITY

Most measurements of the rates of enzyme-catalyzed reactions employ relatively short time periods, conditions that approximate **initial rate conditions**. Under these conditions, only traces of product accumulate, rendering the rate of the reverse reaction negligible. The **initial velocity ( $v_i$ )** of the reaction thus is essentially that of the rate of the forward reaction.

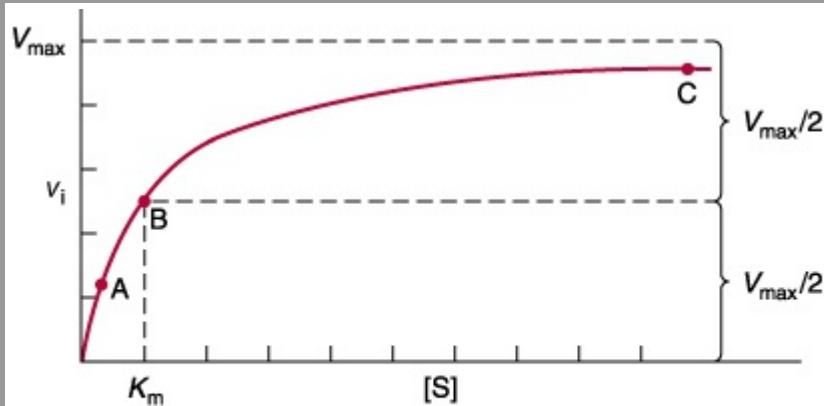
Assays of enzyme activity almost always employ a large ( $10^3$ – $10^7$ ) molar excess of substrate over enzyme. Under these conditions,  $v_i$  is proportionate to the concentration of enzyme. Measuring the initial velocity therefore permits one to estimate the quantity of enzyme present in a biologic sample.

## SUBSTRATE CONCENTRATION AFFECTS THE REACTION RATE

In what follows, enzyme reactions are treated as if they had only a single substrate and a single product. For enzymes with multiple substrates, the principles discussed below apply with equal validity. Moreover, by employing pseudo-first-order conditions (see above), scientists can study the dependence of reaction rate upon an individual reactant through the appropriate choice of fixed and variable substrates. In other words, under pseudo-first-order conditions the behavior of a multisubstrate enzyme will imitate one having a single substrate. In this instance, however, the observed rate constant will be a function of the rate constant  $k_1$  for the reaction as well as the concentration of the fixed substrate(s).

For a typical enzyme, as substrate concentration is increased,  $v_i$  increases until it reaches a maximum value  $V_{max}$  (**Figure 8–4**). When further increases in substrate concentration do no further increase  $v_i$ , the enzyme is said to be "saturated" with the substrate. Note that the shape of the curve that relates activity to substrate concentration (**Figure 8–4**) is hyperbolic. At any given instant, only substrate molecules that are combined with the enzyme as an enzyme-substrate (ES) complex can be transformed into a product. Since the equilibrium constant for the formation of the enzyme-substrate complex is not infinitely large, only a fraction of the enzyme may be present as an ES complex even when the substrate is present in excess (points A and B of **Figure 8–5**). At points A or B, increasing or decreasing [S] therefore will increase or decrease the number of ES complexes with a corresponding change in  $v_i$ . At point C (**Figure 8–5**), however, essentially all the enzyme is present as the ES complex. Since no free enzyme remains available for forming ES, further increases in [S] cannot increase the rate of the reaction. Under these saturating conditions,  $v_i$  depends solely on—and thus is limited by—the rapidity with which product dissociates from the enzyme so that it may combine with more substrate.

**Figure 8–4**

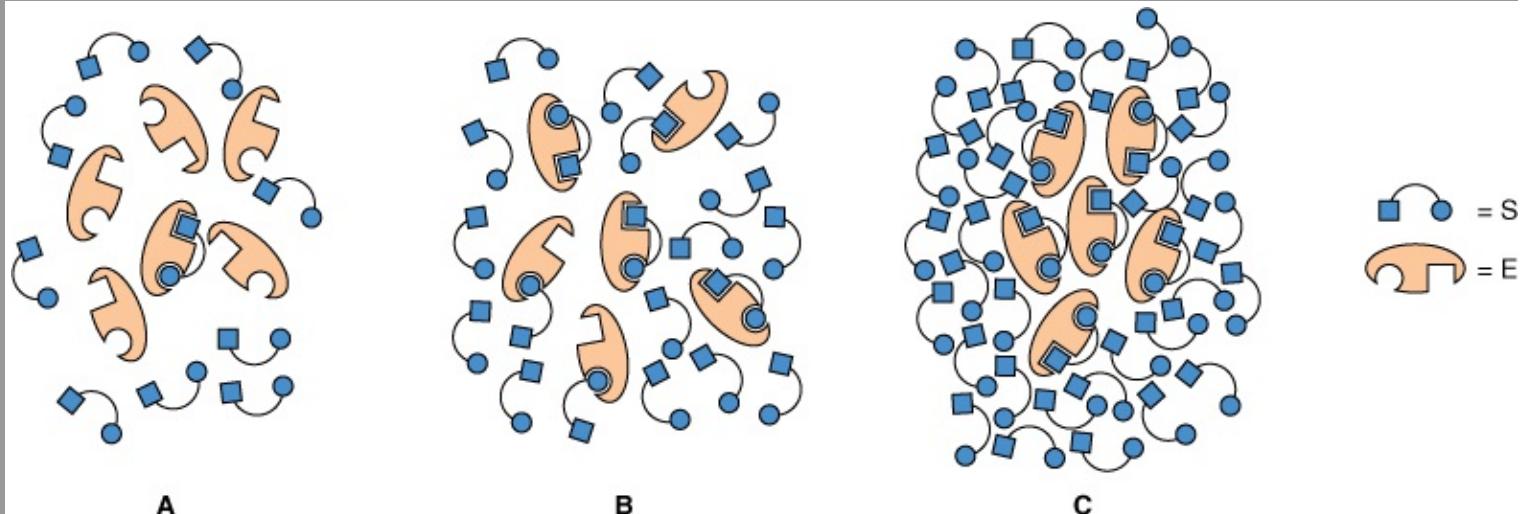


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effect of substrate concentration on the initial velocity of an enzyme-catalyzed reaction.

**Figure 8–5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Representation of an enzyme in the presence of a concentration of substrate that is below  $K_m$  (A), at a concentration equal to  $K_m$  (B), and at a concentration well above  $K_m$  (C). Points A, B, and C correspond to those points in **Figure 8–4**.

## THE MICHAELIS-MENTEN & HILL EQUATIONS MODEL THE EFFECTS OF SUBSTRATE CONCENTRATION

### The Michaelis-Menten Equation

The Michaelis-Menten equation (29) illustrates in mathematical terms the relationship between initial reaction velocity  $v_i$  and substrate concentration  $[S]$ , shown graphically in **Figure 8-4**:

$$v_i = \frac{V_{\max}[S]}{K_m + [S]} \quad (29)$$

**The Michaelis constant  $K_m$  is the substrate concentration at which  $v_i$  is half the maximal velocity ( $V_{\max}/2$ ) attainable at a particular concentration of the enzyme.**  $K_m$  thus has the dimensions of substrate concentration. The dependence of initial reaction velocity on  $[S]$  and  $K_m$  may be illustrated by evaluating the Michaelis-Menten equation under three conditions.

- When  $[S]$  is much less than  $K_m$  (point A in **Figures 8-4** and **8-5**), the term  $K_m + [S]$  is essentially equal to  $K_m$ . Replacing  $K_m + [S]$  with  $K_m$  reduces equation (29) to

$$v_i = \frac{V_{\max}[S]}{K_m + [S]} \quad v_i = \frac{V_{\max}[S]}{K_m} = \left( \frac{V_{\max}}{K_m} \right) [S] \quad (30)$$

where  $\sim$  means "approximately equal to." Since  $V_{\max}$  and  $K_m$  are both constants, their ratio is a constant. In other words, when  $[S]$  is considerably below  $K_m$ ,  $v_i$  is proportionate to  $k[S]$ . The initial reaction velocity therefore is directly proportional to  $[S]$ .

- When  $[S]$  is much greater than  $K_m$  (point C in **Figures 8-4** and **8-5**), the term  $K_m + [S]$  is essentially equal to  $[S]$ . Replacing  $K_m + [S]$  with  $[S]$  reduces equation (29) to

$$v_i = \frac{V_{\max}[S]}{K_m + [S]} \quad v_i = \frac{V_{\max}[S]}{[S]} = V_{\max} \quad (31)$$

Thus, when  $[S]$  greatly exceeds  $K_m$ , the reaction velocity is maximal ( $V_{\max}$ ) and unaffected by further increases in the substrate concentration.

- When  $[S] = K_m$  (point B in **Figures 8-4** and **8-5**):

$$v_i = \frac{V_{\max}[S]}{K_m + [S]} = \frac{V_{\max}[S]}{2[S]} = \frac{V_{\max}}{2} \quad (32)$$

Equation (32) states that when  $[S]$  equals  $K_m$ , the initial velocity is half-maximal. Equation (32) also reveals that  $K_m$  is—and may be determined experimentally from—the substrate concentration at which the initial velocity is half-maximal.

### A Linear Form of the Michaelis-Menten Equation Is Used to Determine $K_m$ & $V_{\max}$

The direct measurement of the numeric value of  $V_{\max}$ , and therefore the calculation of  $K_m$ , often requires impractically high concentrations of substrate to achieve saturating conditions. A linear form of the Michaelis-Menten equation circumvents this difficulty and permits  $V_{\max}$  and  $K_m$  to be extrapolated from initial velocity data obtained at less than saturating concentrations of the substrate. Start with equation (29),

$$v_i = \frac{V_{\max}[S]}{K_m + [S]} \quad (29)$$

invert

$$\frac{1}{v_i} = \frac{K_m + [S]}{V_{\max}[S]} \quad (33)$$

factor

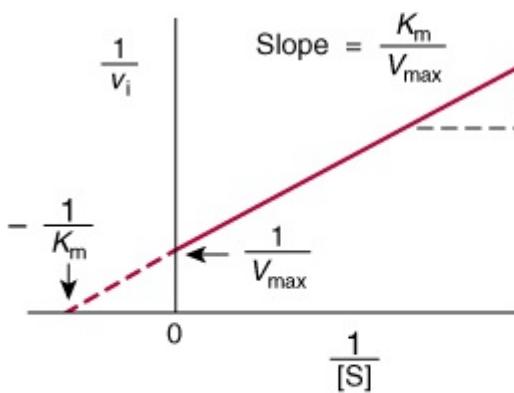
$$\frac{1}{v_i} = \frac{K_m}{V_{\max}[S]} + \frac{[S]}{V_{\max}[S]} \quad (34)$$

and simplify

$$\frac{1}{v_i} = \left( \frac{K_m}{V_{\max}} \right) \frac{1}{[S]} + \frac{1}{V_{\max}} \quad (35)$$

Equation (35) is the equation for a straight line,  $y = ax + b$ , where  $y = 1/v_i$  and  $x = 1/[S]$ . A plot of  $1/v_i$  as  $y$  as a function of  $1/[S]$  as  $x$  therefore gives a straight line whose  $y$  intercept is  $1/V_{\max}$  and whose slope is  $K_m/V_{\max}$ . Such a plot is called a **double reciprocal** or **Lineweaver-Burk plot** (**Figure 8-6**). Setting the  $y$  term of equation (36) equal to zero and solving for  $x$  reveals that the  $x$  intercept is  $-1/K_m$ :

**Figure 8-6**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Double-reciprocal or Lineweaver-Burk plot of  $1/v_i$  versus  $1/[S]$  used to evaluate  $K_m$  and  $V_{max}$ .

$$0 = ax + b; \text{ therefore, } x = \frac{-b}{a} = \frac{-1}{K_m} \quad (36)$$

$K_m$  is thus most readily calculated from the negative  $x$  intercept.

The greatest virtue of the Lineweaver-Burk plot resides in the facility with which it can be used to determine the kinetic mechanisms of enzyme inhibitors (see below). However, in using a double-reciprocal plot to determine kinetic constants it is important to avoid the introduction of bias through the clustering of data at low values of  $1/[S]$ . To avoid this bias, prepare a solution of substrate whose dilution into an assay will produce the maximum desired concentration of the substrate. Now use the same volume of solutions prepared by diluting the stock solution by factors of 1:2, 1:3, 1:4, 1:5, etc. The data will then fall on the  $1/[S]$  axis at intervals of 1, 2, 3, 4, 5, etc. Alternatively, a single-reciprocal plot such as the Eadie-Hofstee ( $v_i$  versus  $v_i/[S]$ ) or Hanes-Woolf ( $[S]/v_i$  versus  $[S]$ ) plot can be used to minimize clustering.

## The Catalytic Constant, $k_{cat}$

Several parameters may be used to compare the relative activity of different enzymes or of different preparations of the same enzyme. The activity of impure enzyme preparations typically is expressed as a *specific activity* ( $V_{max}$  divided by the protein concentration). For a homogeneous enzyme, one may calculate its *turnover number* ( $V_{max}$  divided by the moles of enzyme present). But if the number of active sites present is known, the catalytic activity of a homogeneous enzyme is best expressed as its *catalytic constant*,  $k_{cat}$  ( $V_{max}$  divided by the number of active sites,  $S_t$ ):

$$k_{cat} = \frac{V_{max}}{S_t} \quad (37)$$

Since the units of concentration cancel out, the units of  $k_{cat}$  are reciprocal time.

## Catalytic Efficiency, $k_{cat}/K_m$

By what measure should the efficiency of different enzymes, different substrates for a given enzyme, and the efficiency with which an enzyme catalyzes a reaction in the forward and reverse directions be quantified and compared? While the maximum capacity of a given enzyme to convert substrate to product is important, the benefits of a high  $k_{cat}$  can only be realized if  $K_m$  is sufficiently low. Thus, *catalytic efficiency* of enzymes is best expressed in terms of the ratio of these two kinetic constants,  $k_{cat}/K_m$ .

For certain enzymes, once substrate binds to the active site, it is converted to product and released so rapidly as to render these events effectively instantaneous. For these exceptionally efficient catalysts, the rate-limiting step in catalysis is the formation of the ES complex. Such enzymes are said to be *diffusion-limited*, or catalytically perfect, since the fastest possible rate of catalysis is determined by the rate at which molecules move or diffuse through the solution. Examples of enzymes for which  $k_{cat}/K_m$  approaches the diffusion limit of  $10^8\text{--}10^9 \text{ M}^{-1}\text{s}^{-1}$  include triosephosphate isomerase, carbonic anhydrase, acetylcholinesterase, and adenosine deaminase.

In living cells, the assembly of enzymes that catalyze successive reactions into multimeric complexes can circumvent the limitations imposed by diffusion. The geometric relationships of the enzymes in these complexes are such that the substrates and products do not diffuse into the bulk solution until the last step in the sequence of catalytic steps is complete. Fatty acid synthetase extends this concept one step further by covalently attaching the growing substrate fatty acid chain to a biotin tether that rotates from active site to active site within the complex until synthesis of a palmitic acid molecule is complete (Chapter 23).

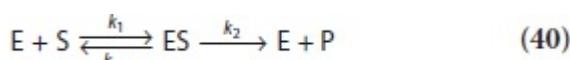
## $K_m$ May Approximate a Binding Constant

The affinity of an enzyme for its substrate is the inverse of the dissociation constant  $K_d$  for dissociation of the enzyme-substrate complex ES:



$$K_d = \frac{k_{-1}}{k_1} \quad (39)$$

Stated another way, the smaller the tendency of the enzyme and its substrate to *dissociate*, the greater the affinity of the enzyme for its substrate. While the Michaelis constant  $K_m$  often approximates the dissociation constant  $K_d$ , this is by no means always the case. For a typical enzyme-catalyzed reaction:



The value of  $[S]$  that gives  $v_i = V_{max}/2$  is

$$[S] = \frac{k_{-1} + k_2}{k_1} = K_m \quad (41)$$

When  $k_{-1} \gg k_2$ , then

$$k_{-1} + k_2 \approx k_{-1} \quad (42)$$

and

$$[S] = \frac{k_1}{k_{-1}} = K_d \quad (43)$$

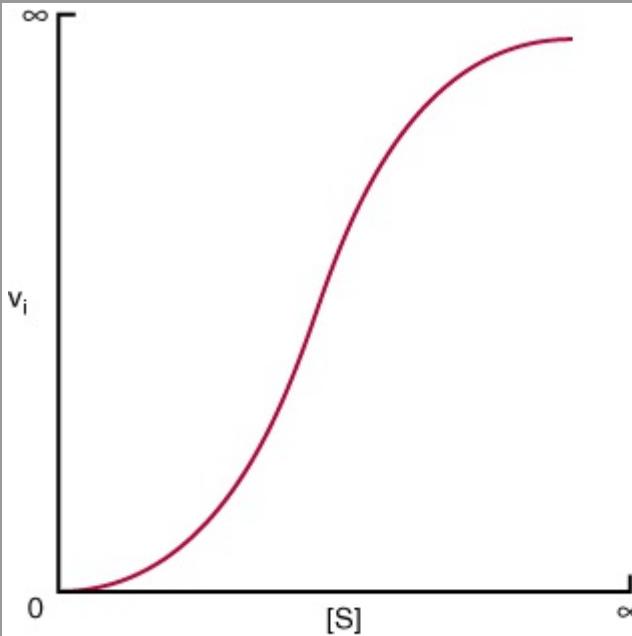
Hence,  $1/K_m$  only approximates  $1/K_d$  under conditions where the association and dissociation of the ES complex are rapid relative to catalysis. For the many enzyme-catalyzed reactions for which  $k_{-1} + k_2$  is not approximately equal to  $k_{-1}$ ,  $1/K_m$  will underestimate  $1/K_d$ .

## The Hill Equation Describes the Behavior of Enzymes that Exhibit Cooperative Binding of Substrate

While most enzymes display the simple **saturation kinetics** depicted in **Figure 8–4** and are adequately described by the Michaelis-Menten expression, some enzymes bind their substrates in a **cooperative** fashion analogous to the binding of oxygen by hemoglobin (Chapter 6). Cooperative behavior is an exclusive property of multimeric enzymes that bind substrate at multiple sites.

For enzymes that display positive cooperativity in binding the substrate, the shape of the curve that relates changes in  $v_i$  to changes in  $[S]$  is sigmoidal (**Figure 8–7**). Neither the Michaelis-Menten expression nor its derived plots can be used to evaluate cooperative kinetics. Enzymologists therefore employ a graphic representation of the **Hill equation** originally derived to describe the cooperative binding of  $O_2$  by hemoglobin. Equation (44) represents the Hill equation arranged in a form that predicts a straight line, where  $k'$  is a complex constant:

**Figure 8–7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

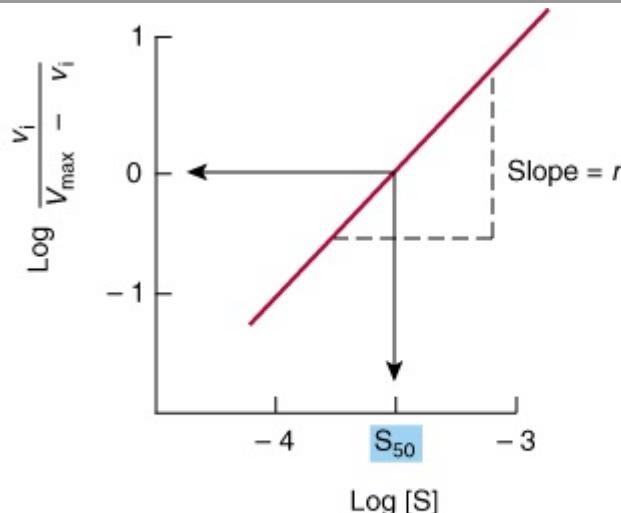
Representation of sigmoid substrate saturation kinetics.

$$\frac{\log v_i}{V_{\max} - v_i} = n \log [S] - \log k' \quad (44)$$

Equation (44) states that when  $[S]$  is low relative to  $k'$ , the initial reaction velocity increases as the  $n$ th power of  $[S]$ .

A graph of  $\log v_i/(V_{\max} - v_i)$  versus  $\log [S]$  gives a straight line (**Figure 8–8**), where the slope of the line  $n$  is the **Hill coefficient**, an empirical parameter whose value is a function of the number, kind, and strength of the interactions of the multiple substrate-binding sites on the enzyme. When  $n = 1$ , all binding sites behave independently and simple Michaelis-Menten kinetic behavior is observed. If  $n$  is greater than 1, the enzyme is said to exhibit positive cooperativity. Binding of substrate to one site then enhances the affinity of the remaining sites to bind additional substrate. The greater the value for  $n$ , the higher the degree of cooperativity and the more markedly sigmoidal will be the plot of  $v_i$  versus  $[S]$ . A perpendicular dropped from the point where the  $y$  term  $\log v_i/(V_{\max} - v_i)$  is zero intersects the  $x$ -axis at a substrate concentration termed  **$S_{50}$** , the substrate concentration that results in half-maximal velocity.  $S_{50}$  thus is analogous to the  $P_{50}$  for oxygen binding to hemoglobin (Chapter 6).

**Figure 8–8**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A graphical representation of a linear form of the Hill equation is used to evaluate  $S_{50}$ , the substrate concentration that produces half-maximal velocity, and the degree of cooperativity  $n$ .



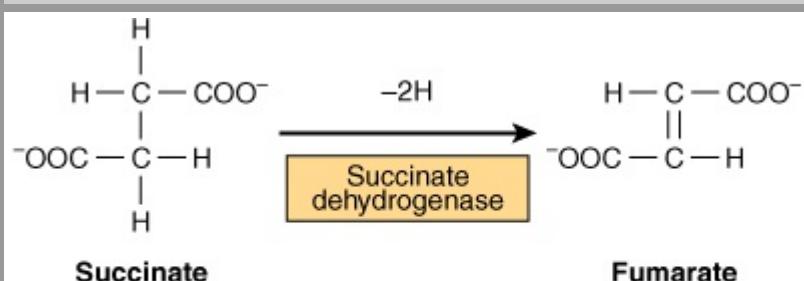
## KINETIC ANALYSIS DISTINGUISHES COMPETITIVE FROM NONCOMPETITIVE INHIBITION

Inhibitors of the catalytic activities of enzymes provide both pharmacologic agents and research tools for the study of the mechanism of enzyme action. The strength of the interaction between an inhibitor and an enzyme depends on forces important in protein structure and ligand binding (hydrogen bonds, electrostatic interactions, hydrophobic interactions, and van der Waals forces; see Chapter 5). Inhibitors can be classified on the basis of their site of action on the enzyme, on whether they chemically modify the enzyme, or on the kinetic parameters they influence. Compounds that mimic the transition state of an enzyme-catalyzed reaction (transition state analogs) or that take advantage of the catalytic machinery of an enzyme (mechanism-based inhibitors) can be particularly potent inhibitors. Kinetically, we distinguish two classes of inhibitors based upon whether raising the substrate concentration does or does not overcome the inhibition.

### Competitive Inhibitors Typically Resemble Substrates

The effects of competitive inhibitors can be overcome by raising the concentration of substrate. Most frequently, in competitive inhibition the inhibitor (**I**) binds to the substrate-binding portion of the active site thereby blocking access by the substrate. The structures of most classic competitive inhibitors therefore tend to resemble the structures of a substrate, and thus are termed **substrate analogs**. Inhibition of the enzyme succinate dehydrogenase by malonate illustrates competitive inhibition by a substrate analog. Succinate dehydrogenase catalyzes the removal of one hydrogen atom from each of the two-methylene carbons of succinate (Figure 8–9). Both succinate and its structural analog malonate ( $\text{HOOC}-\text{CH}_2-\text{COO}^-$ ) can bind to the active site of succinate dehydrogenase, forming an ES or an EI complex, respectively. However, since malonate contains only one methylene carbon, it cannot undergo dehydrogenation.

**Figure 8–9**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The succinate dehydrogenase reaction.

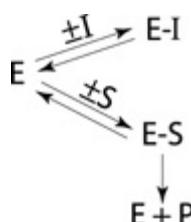
The formation and dissociation of the EI complex is a dynamic process described by



for which the equilibrium constant  $K_i$  is

$$K_i = \frac{[\text{E}][\text{I}]}{[\text{E} - \text{I}]} = \frac{k_1}{k_{-1}} \quad (46)$$

In effect, a **competitive inhibitor acts by decreasing the number of free enzyme molecules available to bind substrate, ie, to form ES, and thus eventually to form product**, as described below.



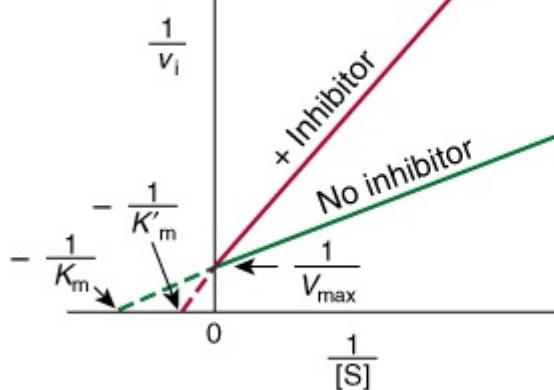
A competitive inhibitor and substrate exert reciprocal effects on the concentration of the EI and ES complexes. Since the formation of ES complexes removes free enzyme available to combine with the inhibitor, increasing [S] decreases the concentration of the EI complex and raises the reaction velocity. The extent to which [S] must be increased to completely overcome the inhibition depends upon the concentration of the inhibitor present, its affinity for the enzyme,  $K_i$ , and the affinity,  $K_m$ , of the enzyme for its substrate.

### Double-Reciprocal Plots Facilitate the Evaluation of Inhibitors

Double-reciprocal plots distinguish between competitive and noncompetitive inhibitors and simplify evaluation of inhibition constants.  $v_i$  is determined at several substrate concentrations both in the presence and in the absence of the inhibitor. For classic competitive inhibition, the lines that connect the experimental data points converge at the y-axis (Figure 8–10). Since the y intercept is equal to  $1/V_{max}$ , this pattern indicates that **when  $1/[S]$  approaches 0,  $v_i$  is independent of the presence of inhibitor**. Note, however, that the intercept on the x-axis does vary with inhibitor concentration—and that since  $-1/K'm$  is smaller than  $1/K_m$ ,  $K'm$  (the "apparent  $K_m$ ") becomes larger in the presence of increasing concentrations of the inhibitor. Thus, a **competitive inhibitor has no effect on  $V_{max}$  but raises  $K'm$ , the apparent  $K_m$  for the substrate**. For a simple competitive inhibition, the intercept on the x-axis is

$$x = \frac{-1}{K_m} \left( 1 + \frac{[I]}{K_i} \right) \quad (47)$$

**Figure 8–10**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Lineweaver-Burk plot of simple competitive inhibition.** Note the complete relief of inhibition at high [S] (ie, low 1/[S]).

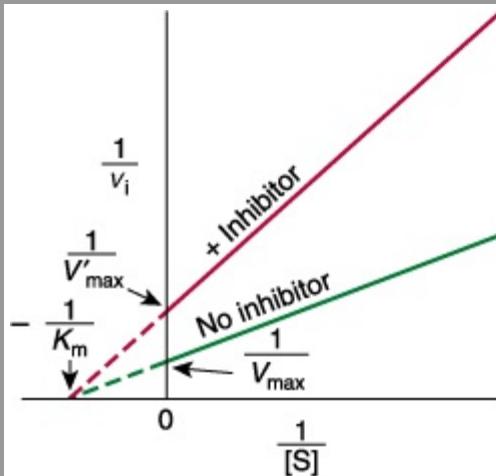
Once  $K_m$  has been determined in the absence of inhibitor,  $K_i$  can be calculated from equation (47).  $K_i$  values are used to compare different inhibitors of the same enzyme. The lower the value for  $K_i$ , the more effective the inhibitor. For example, the statin drugs that act as competitive inhibitors of HMG-CoA reductase (Chapter 26) have  $K_i$  values several orders of magnitude lower than the  $K_m$  for the substrate HMG-CoA.

### Simple Noncompetitive Inhibitors Lower $V_{max}$ But Do Not Affect $K_m$

In strict noncompetitive inhibition, binding of the inhibitor does not affect binding of the substrate. Formation of both EI and EIS complexes is therefore possible. However, while the enzyme-inhibitor complex can still bind the substrate, its efficiency at transforming substrate to product, reflected by  $V_{max}$ , is decreased. Noncompetitive inhibitors bind enzymes at sites distinct from the substrate-binding site and generally bear little or no structural resemblance to the substrate.

For simple noncompetitive inhibition, E and EI possess identical affinity for the substrate, and the EIS complex generates product at a negligible rate (**Figure 8–11**). More complex noncompetitive inhibition occurs when binding of the inhibitor *does* affect the apparent affinity of the enzyme for the substrate, causing the lines to intercept in either the third or fourth quadrants of a double-reciprocal plot (not shown). While certain inhibitors exhibit characteristics of a mixture of competitive and noncompetitive inhibition, the evaluation of these inhibitors exceeds the scope of this chapter.

**Figure 8–11**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

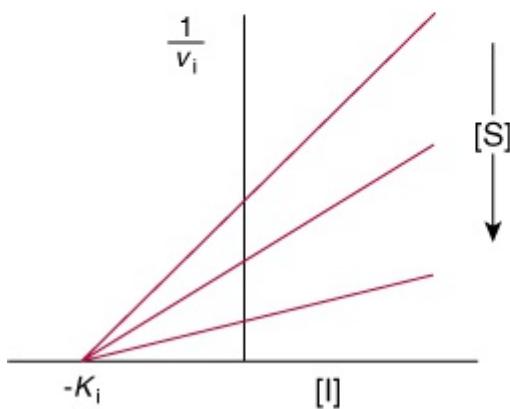
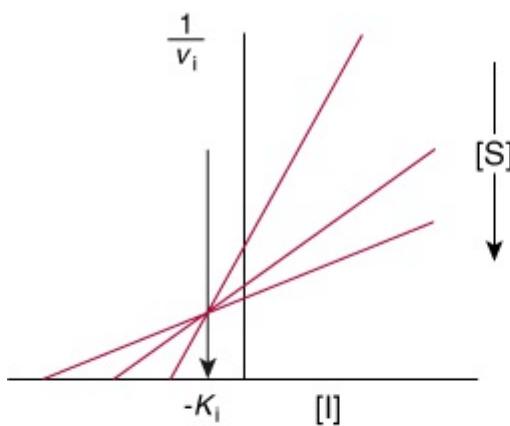
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Lineweaver-Burk plot for simple noncompetitive inhibition.**

### Dixon Plot

A Dixon plot is sometimes employed as an alternative to the Lineweaver-Burk plot for determining inhibition constants. The initial velocity ( $v_i$ ) is measured at several concentrations of inhibitor, but at a fixed concentration of the substrate (S). For a simple competitive or noncompetitive inhibitor, a plot of  $1/v_i$  versus inhibitor concentration [I] yields a straight line. The experiment is repeated at different fixed concentrations of the substrate. The resulting set of lines intersects to the left of the y-axis. For *competitive* inhibition, a perpendicular dropped to the x-axis from the point of intersection of the lines gives  $-K_i$  (**Figure 8–12**, top). For *noncompetitive* inhibition the intercept on the x-axis is  $-K_i$  (**Figure 8–12**, bottom). Pharmaceutical publications frequently employ Dixon plots to illustrate the comparative potency of competitive inhibitors.

**Figure 8–12**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Applications of Dixon plots.** Top: competitive inhibition, estimation of  $K_i$ . Bottom: noncompetitive inhibition, estimation of  $K_i$ .

## IC<sub>50</sub>

A less rigorous alternative to  $K_i$  as a measure of inhibitory potency is the concentration of inhibitor that produces 50% inhibition, **IC<sub>50</sub>**. Unlike the equilibrium dissociation constant  $K_i$ , the numeric value of IC<sub>50</sub> varies as a function of the specific circumstances of substrate concentration, etc, under which it is determined.

### Tightly Bound Inhibitors

Some inhibitors bind to enzymes with such high affinity,  $K_i = 10^{-9}$  M, that the concentration of inhibitor required to measure  $K_i$  falls below the concentration of enzyme typically present in an assay. Under these circumstances, a significant fraction of the total inhibitor may be present as an EI complex. If so, this violates the assumption, implicit in classical steady-state kinetics, that the concentration of free inhibitor is independent of the concentration of enzyme. The kinetic analysis of these tightly bound inhibitors requires specialized kinetic equations that incorporate the concentration of enzyme to estimate  $K_i$  or IC<sub>50</sub> and to distinguish competitive from noncompetitive tightly bound inhibitors.

### Irreversible Inhibitors "Poison" Enzymes

In the above examples, the inhibitors form a dissociable, dynamic complex with the enzyme. Fully active enzyme can therefore be recovered simply by removing the inhibitor from the surrounding medium. However, a variety of other inhibitors act irreversibly by chemically modifying the enzyme. These modifications generally involve making or breaking covalent bonds with aminoacyl residues essential for substrate binding, catalysis, or maintenance of the enzyme's functional conformation. Since these covalent changes are relatively stable, an enzyme that has been "poisoned" by an irreversible inhibitor such as a heavy metal atom or an acylating reagent remains inhibited even after the removal of the remaining inhibitor from the surrounding medium.

### Mechanism-Based Inhibition

"Mechanism-based" or "suicide" inhibitors are specialized substrate analogs that contain a chemical group that can be transformed by the catalytic machinery of the target enzyme. After binding to the active site, catalysis by the enzyme generates a highly reactive group that forms a covalent bond to, and **blocks the function of a catalytically essential residue**. The specificity and persistence of suicide inhibitors, which are both enzyme-specific and unreactive outside the confines of the enzyme's active site, render them promising leads for the development of enzyme-specific drugs. The kinetic analysis of suicide inhibitors lies beyond the scope of this chapter. Neither the Lineweaver-Burk nor the Dixon approach is applicable since suicide inhibitors violate a key boundary condition common to both approaches, namely that the activity of the enzyme does not decrease during the course of the assay.

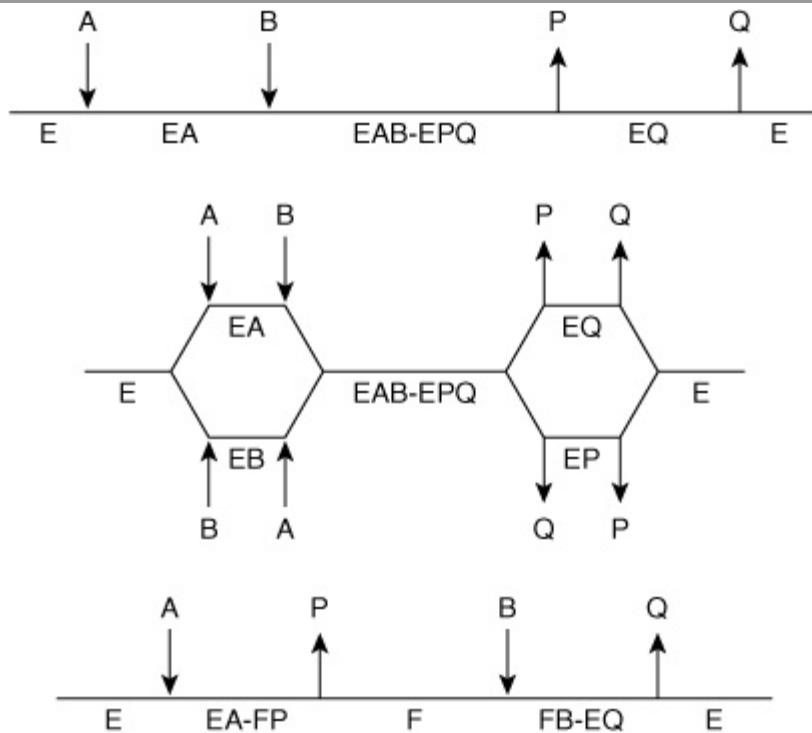
## MOST ENZYME-CATALYZED REACTIONS INVOLVE TWO OR MORE SUBSTRATES

While several enzymes have a single substrate, many others have two—and sometimes more—substrates and products. The fundamental principles discussed above, while illustrated for single-substrate enzymes, apply also to multisubstrate enzymes. The mathematical expressions used to evaluate multisubstrate reactions are, however, complex. While a detailed analysis of the full range of multisubstrate reactions exceeds the scope of this chapter, some common types of kinetic behavior for two-substrate, two-product reactions (termed "BiBi" reactions) are considered below.

### Sequential or Single-Displacement Reactions

In **sequential reactions**, both substrates must combine with the enzyme to form a ternary complex before catalysis can proceed (**Figure 8–13, top**). Sequential reactions are sometimes referred to as single-displacement reactions because the group undergoing transfer is usually passed directly, in a single step, from one substrate to the other. Sequential Bi-Bi reactions can be further distinguished on the basis of whether the two substrates add in a **random** or in a **compulsory** order. For random-order reactions, either substrate A or substrate B may combine first with the enzyme to form an EA or an EB complex (**Figure 8–13, center**). For compulsory-order reactions, A must first combine with E before B can combine with the EA complex. One explanation for why some enzymes employ compulsory-order mechanisms can be found in Koshland's induced fit hypothesis: the addition of A induces a conformational change in the enzyme that aligns residues that recognize and bind B.

**Figure 8–13**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Representations of three classes of Bi–Bi reaction mechanisms.** Horizontal lines represent the enzyme. Arrows indicate the addition of substrates and departure of products. Top: an ordered Bi–Bi reaction, characteristic of many NAD(P)H-dependent oxidoreductases. Center: a random Bi–Bi reaction, characteristic of many kinases and some dehydrogenases. Bottom: a ping–pong reaction, characteristic of aminotransferases and serine proteases.

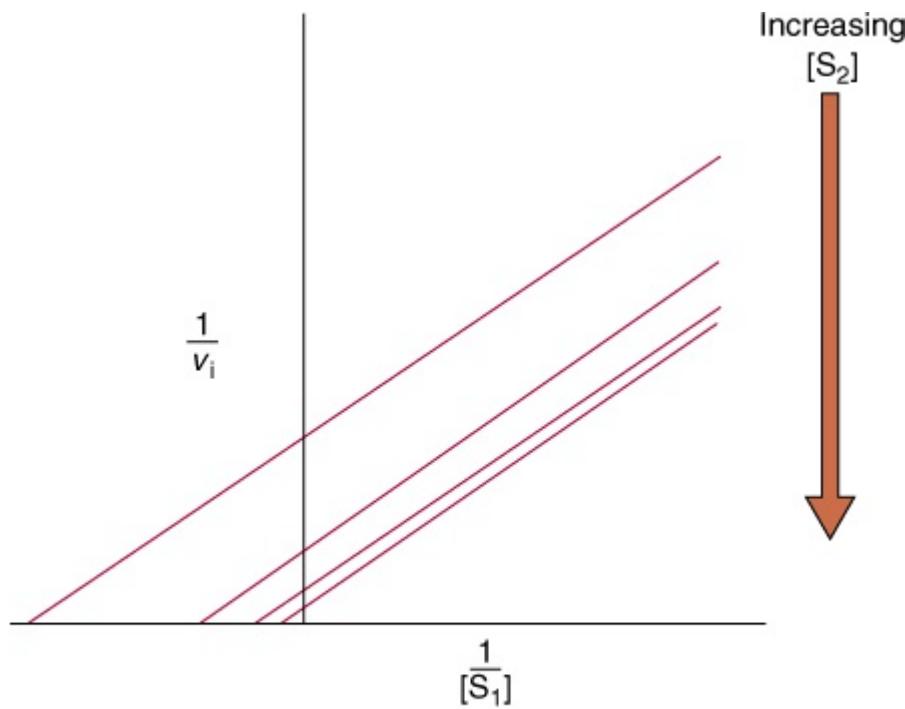
### Ping–Pong Reactions

The term "**ping–pong**" applies to mechanisms in which one or more products are released from the enzyme before all the substrates have been added. Ping-pong reactions involve covalent catalysis and a transient, modified form of the enzyme (see **Figure 7–4**). Ping-pong Bi–Bi reactions are often referred to as **double displacement reactions**. The group undergoing transfer is first displaced from substrate A by the enzyme to form product P and a modified form of the enzyme (F). The subsequent group transfer from F to the second substrate B, forming product Q and regenerating E, constitutes the second displacement (**Figure 8–13, bottom**).

### Most Bi–Bi Reactions Conform to Michaelis–Menten Kinetics

Most Bi–Bi reactions conform to a somewhat more complex form of Michaelis–Menten kinetics in which  $V_{max}$  refers to the reaction rate attained when both substrates are present at saturating levels. Each substrate has its own characteristic  $K_m$  value, which corresponds to the concentration that yields half-maximal velocity when the second substrate is present at saturating levels. As for single-substrate reactions, double-reciprocal plots can be used to determine  $V_{max}$  and  $K_m$ .  $v_i$  is measured as a function of the concentration of one substrate (the variable substrate) while the concentration of the other substrate (the fixed substrate) is maintained constant. If the lines obtained for several fixed-substrate concentrations are plotted on the same graph, it is possible to distinguish a ping-pong mechanism, which yields parallel lines (**Figure 8–14**), from a sequential mechanism, which yields a pattern of intersecting lines (not shown).

**Figure 8–14**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Lineweaver-Burk plot for a two-substrate ping-pong reaction.** An increase in concentration of one substrate ( $S_1$ ) while that of the other substrate ( $S_2$ ) is maintained constant for changes both the  $x$  and  $y$  intercepts, but not the slope.

**Product inhibition studies** are used to complement kinetic analyses and to distinguish between ordered and random Bi-Bi reactions. For example, in a random-order Bi-Bi reaction, each product will act as a competitive inhibitor in the absence of its coproducts regardless of which substrate is designated the variable substrate. However, for a sequential mechanism (Figure 8–13, top), only product Q will give the pattern indicative of competitive inhibition when A is the variable substrate, while only product P will produce this pattern with B as the variable substrate. The other combinations of product inhibitor and variable substrate will produce forms of complex noncompetitive inhibition.

## KNOWLEDGE OF ENZYME KINETICS, MECHANISM, AND INHIBITION AIDS DRUG DEVELOPMENT

### Many Drugs Act as Enzyme Inhibitors

The goal of pharmacology is to identify agents that can

1. Destroy or impair the growth, invasiveness, or development of invading pathogens.
2. Stimulate endogenous defense mechanisms.
3. Halt or impede aberrant molecular processes triggered by genetic, environmental, or biologic stimuli with minimal perturbation of the host's normal cellular functions.

By virtue of their diverse physiologic roles and high degree of substrate selectivity, enzymes constitute natural targets for the development of pharmacologic agents that are both potent and specific. Statin drugs, for example, lower cholesterol production by inhibiting 3-hydroxy-3-methylglutaryl coenzyme A reductase (Chapter 26), while emtricitabine and tenofovir disoproxil fumarate block replication of the human immunodeficiency virus by inhibiting the viral reverse transcriptase (Chapter 34). Pharmacologic treatment of hypertension often includes the administration of an inhibitor of angiotensin-converting enzyme, thus lowering the level of angiotensin II, a vasoconstrictor (Chapter 42).

### Enzyme Kinetics Defines Appropriate Screening Conditions

Enzyme kinetics plays a crucial role in drug discovery. Knowledge of the kinetic behavior of the enzyme of interest is necessary, first and foremost, to select appropriate assay conditions for detecting the presence of an inhibitor. The concentration of substrate, for example, must be adjusted such that sufficient product is generated to permit facile detection of the enzyme's activity without being so high that it masks the presence of an inhibitor. Second, enzyme kinetics provides the means for quantifying and comparing the potency of different inhibitors and defining their mode of action. Noncompetitive inhibitors are particularly desirable, because—by contrast to competitive inhibitors—their effects can never be completely overcome by increases in substrate concentration.

### Most Drugs Are Metabolized In Vivo

Drug development often involves more than the kinetic evaluation of the interaction of inhibitors with the target enzyme. Drugs may be acted upon by enzymes present in the patient or pathogen, a process termed **drug metabolism**. For example, penicillin and other  $\beta$ -lactam antibiotics block cell wall synthesis in bacteria by irreversibly poisoning the enzyme alanyl alanine carboxypeptidase-transpeptidase. Many bacteria, however, produce  $\beta$ -lactamases that hydrolyze the critical  $\beta$ -lactam function in penicillin and related drugs. One strategy for overcoming the resulting antibiotic resistance is to simultaneously administer a  $\beta$ -lactamase inhibitor with a  $\beta$ -lactam antibiotic.

Metabolic transformation is sometimes required to convert an inactive drug precursor, or **prodrug**, into its biologically active form (Chapter 53). 2'-Deoxy-5-fluorouridylic acid, a potent inhibitor of thymidylate synthase, a common target of cancer chemotherapy, is produced from 5-fluorouracil via a series of enzymatic transformations catalyzed by a phosphoribosyl transferase and the enzymes of the deoxyribonucleoside salvage pathway (Chapter 33). The effective design and administration of prodrugs requires knowledge of the kinetics and mechanisms of the enzymes responsible for transforming them into their biologically active forms.

## SUMMARY

- The study of enzyme kinetics—the factors that affect the rates of enzyme-catalyzed reactions—reveals the individual steps by which enzymes transform substrates into products.
- $\Delta G$ , the overall change in free energy for a reaction, is independent of reaction mechanism and provides no information concerning *rates* of reactions.
- $K_{eq}$ , a ratio of reaction *rate constants*, may be calculated from the concentrations of substrates and products at equilibrium or from the ratio  $k_1/k_{-1}$ . Enzymes do not affect  $K_{eq}$ .
- Reactions proceed via transition states for which  $\Delta G_F$  is the activation energy. Temperature, hydrogen ion concentration, enzyme concentration, substrate concentration, and inhibitors all affect the rates of enzyme-catalyzed reactions.
- Measurement of the rate of an enzyme-catalyzed reaction generally employs initial rate conditions, for which the virtual absence of product precludes the reverse reaction.
- Linear forms of the Michaelis-Menten equation simplify determination of  $K_m$  and  $V_{max}$ .
- A linear form of the Hill equation is used to evaluate the cooperative substrate-binding kinetics exhibited by some multimeric enzymes. The slope  $n$ , the Hill coefficient, reflects the number, nature, and strength of the interactions of the substrate-binding sites. A value of  $n$  greater than 1 indicates positive cooperativity.
- The effects of simple competitive inhibitors, which typically resemble substrates, are overcome by raising the concentration of the substrate. Simple noncompetitive inhibitors lower  $V_{max}$  but do not affect  $K_m$ .
- For simple competitive and noncompetitive inhibitors, the inhibitory constant  $K_i$  is equal to the equilibrium dissociation constant for the relevant enzyme-inhibitor complex. A simpler and less rigorous term for evaluating the effectiveness of an inhibitor is  $IC_{50}$ , the concentration of inhibitor that produces 50% inhibition under the particular circumstances of the experiment.
- Substrates may add in a random order (either substrate may combine first with the enzyme) or in a compulsory order (substrate A must bind before substrate B).
- In ping-pong reactions, one or more products are released from the enzyme before all the substrates have been added.
- Applied enzyme kinetics facilitates the identification and characterization of drugs that selectively inhibit specific enzymes. Enzyme kinetics thus plays a central and critical role in drug discovery, in comparative pharmacodynamics, and in determining the mode of action of drugs.

## REFERENCES

- Cook PF, Cleland WW: *Enzyme Kinetics and Mechanism*. Garland Science, 2007.
- Copeland RA: *Evaluation of Enzyme Inhibitors in Drug Discovery*. John Wiley & Sons, 2005.
- Cornish-Bowden A: *Fundamentals of Enzyme Kinetics*. Portland Press Ltd, 2004.
- Dixon M: The determination of enzyme inhibitor constants. *Biochem J* 1953;55:170.
- Dixon M: The graphical determination of  $K_m$  and  $K_i$ . *Biochem J* 1972;129:197.
- Fersht A: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Freeman, 1999.
- Fraser CM, Rappuoli R: Application of microbial genomic science to advanced therapeutics. *Annu Rev Med* 2005;56:459.
- Henderson PJF: A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors. *Biochem J* 1972;127:321.
- Schramm, VL: Enzymatic transition-state theory and transition-state analogue design. *J Biol Chem* 2007;282:28297.
- Schultz AR: *Enzyme Kinetics: From Diastase to Multi-enzyme Systems*. Cambridge University Press, 1994.
- Segel IH: *Enzyme Kinetics*. Wiley Interscience, 1975.
- Wlodawer A: Rational approach to AIDS drug design through structural biology. *Annu Rev Med* 2002;53:595.

## OBJECTIVES

After studying this chapter, you should be able to:

- Explain the concept of whole-body homeostasis and its response to fluctuations in the external environment.
- Discuss why the cellular concentrations of substrates for most enzymes tend to be close to  $K_m$ .
- List multiple mechanisms by which active control of metabolite flux is achieved.
- Describe the advantages of certain enzymes being elaborated as proenzymes.
- Illustrate the physiologic events that trigger the conversion of a proenzyme to the corresponding active enzyme.
- Describe typical structural changes that accompany conversion of a proenzyme to the active enzyme.
- Describe the basic features of a typical binding site for metabolites and second messengers that regulate catalytic activity of certain enzymes.
- Indicate two general ways in which an allosteric effector can modify catalytic activity.
- Outline the roles of protein kinases, protein phosphatases, and of regulatory and hormonal and second messengers in initiating a metabolic process.

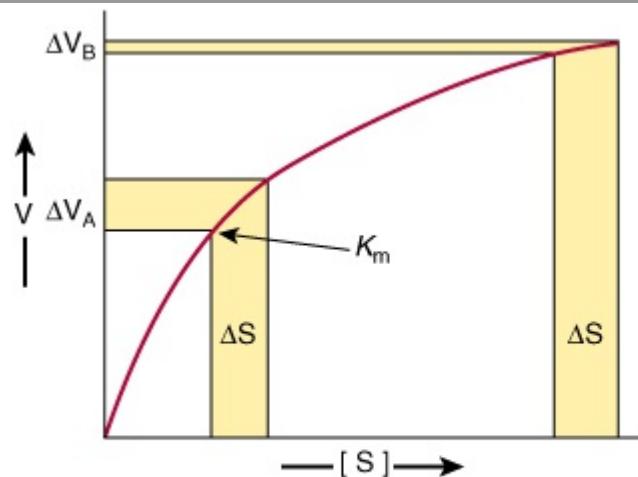
## BIO MEDICAL IMPORTANCE

The nineteenth-century physiologist Claude Bernard enunciated the conceptual basis for metabolic regulation. He observed that living organisms respond in ways that are both quantitatively and temporally appropriate to permit them to survive the multiple challenges posed by changes in their external and internal environments. Walter Cannon subsequently coined the term "homeostasis" to describe the ability of animals to maintain a constant intracellular environment despite changes in their external environment. We now know that organisms respond to changes in their external and internal environment by balanced, coordinated adjustments in the rates of specific metabolic reactions. Perturbations of the sensor-response machinery responsible for maintaining homeostatic balance can be deleterious to human health. Cancer, diabetes, cystic fibrosis, and Alzheimer's disease, for example, are all characterized by regulatory dysfunctions triggered by pathogenic agents or genetic mutations. Many oncogenic viruses elaborate protein-tyrosine kinases that modify the regulatory events that control patterns of gene expression, contributing to the initiation and progression of cancer. The toxin from *Vibrio cholerae*, the causative agent of cholera, disables sensor-response pathways in intestinal epithelial cells by ADP-ribosylating the GTP-binding proteins (G-proteins) that link cell surface receptors to adenylyl cyclase. The consequent activation of the cyclase leads to the unrestricted flow of water into the intestines, resulting in massive diarrhea and dehydration. *Yersinia pestis*, the causative agent of plague, elaborates a protein-tyrosine phosphatase that hydrolyzes phosphoryl groups on key cytoskeletal proteins. Dysfunctions in the proteolytic systems responsible for the degradation of defective or abnormal proteins are believed to play a role in neurodegenerative diseases such as Alzheimer and Parkinson's. In addition to their immediate function as regulators of enzyme activity, protein degradation, etc., covalent modifications such as phosphorylation, acetylation, and ubiquitination provide a protein-based code for the storage and hereditary transmission of information (Chapter 35). Such DNA-independent information systems are referred to as **epigenetic**. Knowledge of factors that control the rates of enzyme-catalyzed reactions thus is essential to an understanding of the molecular basis of disease and its transmission. This chapter outlines the patterns by which metabolic processes are controlled, and provides illustrative examples. Subsequent chapters provide additional examples.

## REGULATION OF METABOLITE FLOW CAN BE ACTIVE OR PASSIVE

Enzymes that operate at their maximal rate cannot respond to increases in substrate concentration, and can respond only to precipitous decreases in substrate concentration. The  $K_m$  values for most enzymes, therefore, tend to be close to the average intracellular concentration of their substrates, so that changes in substrate concentration generate corresponding changes in the metabolite flux (Figure 9-1). Responses to changes in substrate level represent an important but *passive* means for coordinating metabolite flow and maintaining homeostasis in quiescent cells. However, they offer a limited scope for responding to changes in environmental variables. The mechanisms that regulate enzyme efficiency in an *active* manner in response to internal and external signals are discussed below.

**Figure 9-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

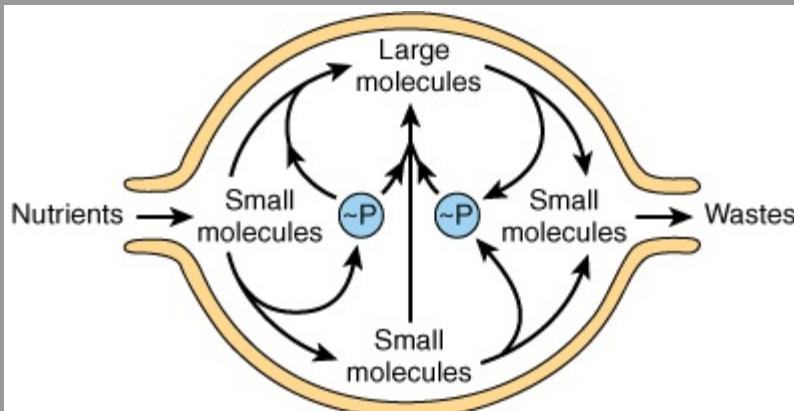
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Differential response of the rate of an enzyme-catalyzed reaction,  $\frac{dV}{d[S]}$ , to the same incremental change in substrate concentration at a substrate concentration close to  $K_m$  ( $\Delta V_A$ ) or far above  $K_m$  ( $\Delta V_B$ ).

## Metabolite Flow Tends to Be Unidirectional

Despite the existence of short-term oscillations in metabolite concentrations and enzyme levels, living cells exist in a dynamic steady state in which the mean concentrations of metabolic intermediates remain relatively constant over time. While all chemical reactions are to some extent reversible, in living cells the reaction products serve as substrates for—and are removed by—other enzyme-catalyzed reactions (Figure 9-2). Many nominally reversible reactions thus occur unidirectionally. This succession of coupled metabolic reactions is accompanied by an overall change in free energy that favors unidirectional metabolite flow (Chapter 11). The unidirectional flow of metabolites through a pathway with a large overall negative change in free energy is analogous to the flow of water through a pipe in which one end is lower than the other. Bends or kinks in the pipe simulate individual enzyme-catalyzed steps with a small negative or positive change in free energy. Flow of water through the pipe nevertheless remains unidirectional due to the overall change in height, which corresponds to the overall change in free energy in a pathway (Figure 9-3).

**Figure 9-2**

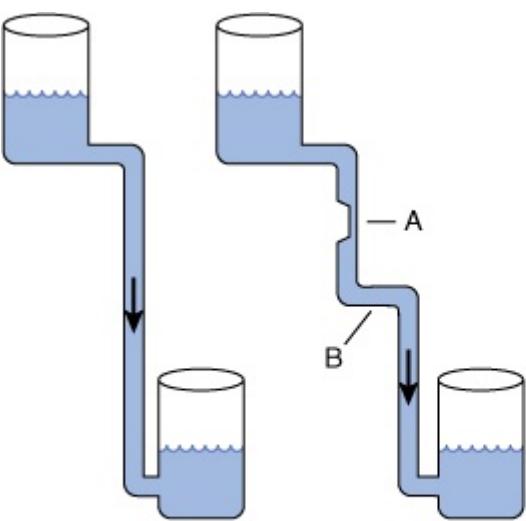


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An idealized cell in steady state. Note that metabolite flow is unidirectional.

**Figure 9-3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hydrostatic analogy for a pathway with a rate-limiting step (A) and a step with a  $\Delta G$  value near 0 (B).

## COMPARTMENTATION ENSURES METABOLIC EFFICIENCY & SIMPLIFIES REGULATION

In eukaryotes, anabolic and catabolic pathways that interconvert common products may take place in specific subcellular compartments. For example, many of the enzymes that degrade proteins and polysaccharides reside inside organelles called lysosomes. Similarly, fatty acid biosynthesis occurs in the cytosol, whereas fatty acid oxidation takes place within mitochondria (Chapters 22 and 23). Segregation of certain metabolic pathways within specialized cell types provides a further means for physical compartmentation.

Fortunately, many apparently antagonistic pathways can coexist in the absence of physical barriers, provided that thermodynamics dictates that each proceeds with the formation of one or more *unique intermediates*. For any reaction or series of reactions, the change in free energy that takes place when metabolite flow proceeds in the "forward" direction is equal in magnitude *but opposite in sign* from that required to proceed in the reverse direction. Some enzymes within these pathways catalyze reactions, such as isomerizations, that can act as bidirectional catalysts *in vivo* because the difference in free energy between substrates and products is close to zero. However, they represent the exception rather than the rule. Virtually all metabolic pathways proceed via one or more steps for which  $\Delta G$  is significant. For example glycolysis, the breakdown of glucose to form two molecules of pyruvate, has a favorable overall  $\Delta G$  of  $-96 \text{ kJ/mol}$ , a value much too large to simply operate in "reverse" when wishing to convert excess pyruvate to glucose. Consequently, gluconeogenesis proceeds via a pathway in which the three most energetically disfavored steps from glycolysis are replaced by new reactions catalyzed by distinct enzymes (Chapter 20).

The ability of enzymes to discriminate between the structurally similar coenzymes  $\text{NAD}^+$  and  $\text{NADP}^+$  also results in a form of compartmentation. The reduced forms of both coenzymes are not readily distinguishable. However, the reactions that generate and later consume electrons that are destined for ATP generation are segregated in NADH, away from those used in the reductive steps of many biosynthetic pathways, which are carried by NADPH.

### Controlling an Enzyme that Catalyzes a Rate-Limiting Reaction Regulates an Entire Metabolic Pathway

While the flux of metabolites through metabolic pathways involves catalysis by numerous enzymes, active control of homeostasis is achieved by the regulation of only a select subset of these enzymes. The ideal enzyme for regulatory intervention is one whose quantity or catalytic efficiency dictates that the reaction it catalyzes is slow relative to all others in the pathway. Decreasing the catalytic efficiency or the quantity of the catalyst responsible for the "bottleneck" or **rate-limiting reaction** immediately reduces metabolite flux through the entire pathway. Conversely, an increase in either its quantity or catalytic efficiency enhances flux through the pathway as a whole. For example, acetyl-CoA carboxylase catalyzes the synthesis of malonyl-CoA, the first committed reaction of fatty acid biosynthesis (Chapter 23). When synthesis of malonyl-CoA is inhibited, subsequent reactions of fatty acid synthesis cease for lack of substrates. As natural "governors" of metabolic flux, the enzymes that catalyze rate-limiting steps also constitute efficient targets for regulatory intervention by drugs. For example, "statin" drugs curtail synthesis of cholesterol by inhibiting HMG-CoA reductase, which catalyzes the rate-limiting reaction of cholesterologenesis.

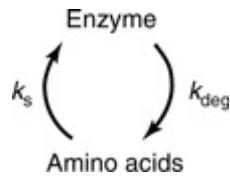
## REGULATION OF ENZYME QUANTITY

The catalytic capacity of the rate-limiting reaction in a metabolic pathway is the product of the concentration of enzyme molecules and their intrinsic catalytic efficiency. It therefore follows that catalytic capacity can be influenced both by changing the quantity of enzyme present and by altering its intrinsic catalytic efficiency.

### Proteins Are Continuously Synthesized and Degraded

By measuring the rates of incorporation of  $^{15}\text{N}$ -labeled amino acids into protein and the rates of loss of  $^{15}\text{N}$  from protein, Schoenheimer deduced that body proteins are in a state of "dynamic equilibrium" in which they are continuously synthesized and degraded—a process referred to as **protein turnover**. This holds even for those proteins that are present at an essentially constant, or **constitutive**, steady-state level over time. On the other hand, the concentrations of many enzymes are influenced by a wide range of physiologic, hormonal, or dietary factors.

The absolute quantity of an enzyme reflects the net balance between its rate of synthesis and its rate of degradation. In human subjects, alterations in the levels of specific enzymes can be effected by a change in the rate constant for the overall processes of synthesis ( $k_s$ ), degradation ( $k_{\text{deg}}$ ), or both.



### Control of Enzyme Synthesis

The synthesis of certain enzymes depends upon the presence of **inducers**, typically substrates or structurally related compounds that stimulate the transcription of the gene that encodes them (Chapters 36 and 37). *Escherichia coli* grown on glucose will, for example, only catabolize lactose after addition of a  $\beta$ -galactoside, an inducer that triggers synthesis of a

$\beta$ -galactosidase and a galactoside permease (Figure 38–3). Inducible enzymes of humans include tryptophan pyrolase, threonine dehydratase, tyrosine– $\alpha$ -ketoglutarate aminotransferase, enzymes of the urea cycle, HMG-CoA reductase, and cytochrome P450. Conversely, an excess of a metabolite may curtail synthesis of its cognate enzyme via **repression**. Both induction and repression involve *cis* elements, specific DNA sequences located upstream of regulated genes, and trans-acting regulatory proteins. The molecular mechanisms of induction and repression are discussed in Chapter 38. The synthesis of other enzymes can be stimulated by the interaction of hormones and other extracellular signals with specific cell-surface receptors. Detailed information on the control of protein synthesis in response to hormonal stimuli can be found in Chapter 42.

### Control of Enzyme Degradation

In animals many proteins are degraded by the ubiquitin-proteasome pathway, the discovery of which earned Aaron Ciechanover, Avram Hershko, and Irwin Rose a Nobel Prize. Degradation takes place in the 26S proteasome, a large macromolecular complex made up of more than 30 polypeptide subunits arranged in the form of a hollow cylinder. The active sites of its proteolytic subunits face the interior of the cylinder, thus preventing indiscriminate degradation of cellular proteins. Proteins are targeted to the interior of the proteasome by "ubiquitination," the covalent attachment of one or more ubiquitin molecules. Ubiquitin is a small, approximately 75 residue, protein that is highly conserved among eukaryotes.

Ubiquitin-proteasome is catalyzed by a large family of enzymes called E3 ligases, which attach ubiquitin to the side-chain amino group of lysyl residues. The ubiquitin-proteasome pathway is responsible both for the regulated degradation of selected cellular proteins (for example, cyclins—Chapter 35) and for the removal of defective or aberrant protein species. The key to the versatility and selectivity of the ubiquitin-proteasome system resides in both the variety of intracellular E3 ligases and their ability to discriminate between the different physical or conformational states of target proteins. Thus, the ubiquitin-proteasome pathway can selectively degrade proteins whose physical integrity and functional competency have been compromised by the loss of or damage to a prosthetic group, oxidation of cysteine or histidine residues, or deamidation of asparagine or glutamine residues. Recognition by proteolytic enzymes also can be regulated by covalent modifications such as phosphorylation; binding of substrates or allosteric effectors; or association with membranes, oligonucleotides, or other proteins. A growing body of evidence suggests that dysfunctions of the ubiquitin-proteasome pathway contribute to the accumulation of aberrantly folded protein species characteristic of several neurodegenerative diseases.

## MULTIPLE OPTIONS ARE AVAILABLE FOR REGULATING CATALYTIC ACTIVITY

In humans the induction of protein synthesis is a complex multistep process that typically requires hours to produce significant changes in overall enzyme level. By contrast, changes in intrinsic catalytic efficiency effected by binding of dissociable ligands (**allosteric regulation**) or by **covalent modification** achieve regulation of enzymic activity within seconds. Consequently, changes in protein level generally dominate when meeting long-term adaptive requirements, whereas changes in catalytic efficiency are best suited for rapid and transient alterations in metabolite flux.

## ALLOSTERIC EFFECTORS REGULATE CERTAIN ENZYMES

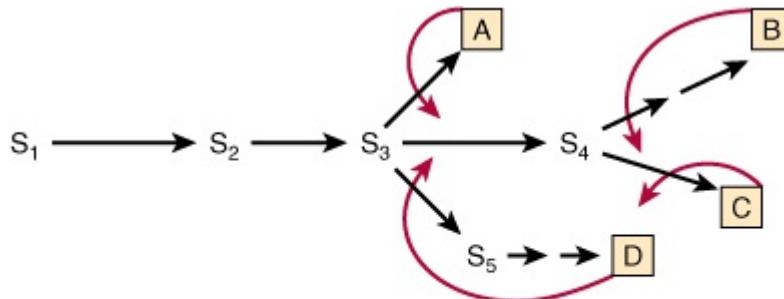
Feedback inhibition refers to the process by which the end product of a multistep biosynthetic pathway binds to and inhibits an enzyme catalyzing one of the early steps in that pathway. In the following example, for the biosynthesis of D from A catalyzed by enzymes Enz<sub>1</sub> through Enz<sub>3</sub>:

Enz<sub>1</sub>    Enz<sub>2</sub>    Enz<sub>3</sub>

A → B → C → D

high concentrations of D inhibit the conversion of A to B. In this example, the feedback inhibitor D acts as a **negative allosteric effector** of Enz<sub>1</sub>. Inhibition results, not from the "backing up" of intermediates, but from the ability of D to bind to and inhibit Enz<sub>1</sub>. Generally, D binds at an **allosteric site**, one spatially distinct from the catalytic site of the target enzyme. Feedback inhibitors thus typically bear little or no structural similarity to the substrates of the enzymes they inhibit. For example, NAD<sup>+</sup> and 3-phosphoglycerate, the substrates for 3-phosphoglycerate dehydrogenase, which catalyzes the first committed step in serine biosynthesis, bear no resemblance to the feedback inhibitor serine. In branched biosynthetic pathways, such as those responsible for nucleotide biosynthesis (Chapter 33), the initial reactions supply intermediates required for the synthesis of multiple end products. **Figure 9-4** shows a hypothetical branched biosynthetic pathway in which curved arrows lead from feedback inhibitors to the enzymes whose activity they inhibit. The sequences S<sub>3</sub> → A, S<sub>4</sub> → B, S<sub>4</sub> → C, and S<sub>3</sub> → D each represent linear reaction sequences that are feedback-inhibited by their end products. Branch point enzymes thus can be targeted to route metabolite flow.

**Figure 9-4**



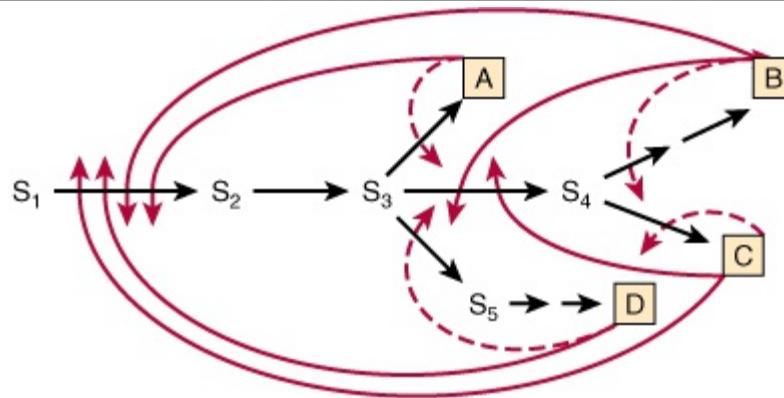
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Sites of feedback inhibition in a branched biosynthetic pathway.** S<sub>1</sub>-S<sub>5</sub> are intermediates in the biosynthesis of end products A-D. Straight arrows represent enzymes catalyzing the indicated conversions. Curved red arrows represent feedback loops and indicate sites of feedback inhibition by specific end products.

Feedback inhibitors typically inhibit the first committed step in a particular biosynthetic sequence. The kinetics of feedback inhibition may be competitive, noncompetitive, partially competitive, or mixed. Layering multiple feedback loops can provide additional fine control. For example, as shown in **Figure 9-5**, the presence of excess product B decreases the requirement for substrate S<sub>2</sub>. However, S<sub>2</sub> is also required for synthesis of A, C, and D. Therefore, for this pathway, excess B curtails synthesis of all four end products, regardless of the need for the other three. To circumvent this potential difficulty, each end product may only partially inhibit catalytic activity. The effect of an excess of two or more end products may be strictly additive or, alternatively, greater than their individual effect (cooperative feedback inhibition).

**Figure 9-5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Multiple feedback inhibition in a branched biosynthetic pathway.** Superimposed on simple feedback loops (dashed red arrows) are multiple feedback loops (solid red arrows) that regulate enzymes common to biosynthesis of several end products.

## Aspartate Transcarbamoylase Is a Model Allosteric Enzyme

Aspartate transcarbamoylase (ATCase), the catalyst for the first reaction unique to pyrimidine biosynthesis (**Figure 33-9**), is a target of feedback regulation by two nucleotide triphosphates: cytidine triphosphate (CTP) and adenosine triphosphate. CTP, an end product of the pyrimidine biosynthetic pathway, inhibits ATCase, whereas the purine nucleotide ATP activates it. Moreover, high levels of ATP can overcome inhibition by CTP, enabling synthesis of *pyrimidine* nucleotides to proceed when *purine* nucleotide levels are elevated.

## Allosteric & Catalytic Sites Are Spatially Distinct

Jacques Monod proposed the existence of allosteric sites that are physically distinct from the catalytic site. He reasoned that the lack of structural similarity between a feedback inhibitor and the substrate(s) for the enzyme whose activity it regulates indicated that these effectors are not **isosteric** with a substrate but **allosteric** ("occupy another space"). **Allosteric enzymes** thus are those for which catalysis at the active site may be modulated by the presence of effectors at an allosteric site. The existence of spatially distinct active and allosteric sites has since been verified in several enzymes using many lines of evidence. For example, x-ray crystallography revealed that the ATCase of *E. coli* consists of six catalytic subunits and six regulatory subunits, the latter of which bind the nucleotide triphosphates that modulate activity. In general, binding of an allosteric regulator induces a conformational change in the enzyme that encompasses the active site.

## Allosteric Effects May Be on K<sub>m</sub> or on V<sub>max</sub>

To refer to the kinetics of allosteric inhibition as "competitive" or "noncompetitive" with substrate carries misleading mechanistic implications. We refer instead to two classes of allosterically regulated enzymes: K-series and V-series enzymes. For K-series allosteric enzymes, the substrate saturation kinetics is competitive in the sense that  $K_m$  is raised without an effect on  $V_{max}$ . For V-series allosteric enzymes, the allosteric inhibitor lowers  $V_{max}$  without affecting the  $K_m$ . Alterations in  $K_m$  or  $V_{max}$  often are the product of conformational changes at the catalytic site induced by binding of the allosteric effector at its site. For a K-series allosteric enzyme, this conformational change may weaken the bonds between substrate and substrate-binding residues. For a V-series allosteric enzyme, the primary effect may be to alter the orientation or charge of catalytic residues, lowering  $V_{max}$ . Intermediate effects on  $K_m$  and  $V_{max}$ , however, may be observed consequent to these conformational changes.

## FEEDBACK REGULATION IS NOT SYNONYMOUS WITH FEEDBACK INHIBITION

In both mammalian and bacterial cells, some end products "feed back" to control their own synthesis, in many instances by feedback inhibition of an early biosynthetic enzyme. We must, however, distinguish between **feedback regulation**, a phenomenologic term devoid of mechanistic implications, and **feedback inhibition**, a mechanism for regulation of enzyme activity. For example, while dietary cholesterol decreases hepatic synthesis of cholesterol, this feedback **regulation** does not involve feedback **inhibition**. HMG-CoA reductase, the rate-limiting enzyme of cholesterologenesis, is affected, but cholesterol does not inhibit its activity. Rather, regulation in response to dietary cholesterol involves curtailment by cholesterol or a cholesterol metabolite of the expression of the gene that encodes HMG-CoA reductase (enzyme repression) (Chapter 26).

## MANY HORMONES ACT THROUGH ALLOSTERIC SECOND MESSENGERS

Nerve impulses and the binding of many hormones to cell surface receptors elicit changes in the rate of enzyme-catalyzed reactions within target cells by inducing the release or synthesis of specialized allosteric effectors called **second messengers**. The primary, or "first," messenger is the hormone molecule or nerve impulse. Second messengers include 3', 5'-cAMP, synthesized from ATP by the enzyme adenylyl cyclase in response to the hormone epinephrine, and  $\text{Ca}^{2+}$ , which is stored inside the endoplasmic reticulum of most cells. Membrane depolarization resulting from a nerve impulse opens a membrane channel that releases calcium ions into the cytoplasm, where they bind to and activate enzymes involved in the regulation of muscle contraction and the mobilization of stored glucose from glycogen. Glucose then supplies the increased energy demands of muscle contraction. Other second messengers include 3',5'-cGMP, nitric oxide, and the polyphosphoinositols produced by the hydrolysis of inositol phospholipids by hormone-regulated phospholipases. Specific examples of the participation of second messengers in the regulation of cellular processes can be found in Chapters 19, 42, and 48.

## REGULATORY COVALENT MODIFICATIONS CAN BE REVERSIBLE OR IRREVERSIBLE

In mammalian cells, a wide range of regulatory covalent modifications occur. **Partial proteolysis** and **phosphorylation**, for example, are frequently employed to regulate the catalytic activity of enzymes. On the other hand, histones and other DNA binding proteins in chromatin are subject to extensive modification by **acetylation**, **methylation**, **ADP-ribosylation**, as well as phosphorylation. The latter modifications, which modulate the manner in which the proteins within chromatin interact with each other as well as the DNA itself, constitute the basis for the "histone code." The resulting changes in chromatin structure within the region affected can render genes more accessible to the protein responsible for their transcription, thereby enhancing gene expression or, on a larger scale, facilitating replication of the entire genome (Chapter 38). On the other hand, changes in chromatin structure that restrict the accessibility of genes to transcription factors, DNA-dependent RNA polymerases, etc, thereby inhibiting transcription, are said to **silence** gene expression.

The histone code represents a classic example of **epigenetics**, the hereditary transmission of information by a means other than the sequence of nucleotides that comprise the genome. In this instance, the pattern of gene expression within a newly formed "daughter" cell will be determined, in part, by the particular set of histone covalent modifications embodied in the chromatin proteins inherited from the "parental" cell.

Acetylation, ADP-ribosylation, methylation, and phosphorylation are all examples of "reversible" covalent modifications. In this instance, reversible refers to the fact that the modified protein can be restored to its original, modification-free state. It does not, however, refer to the mechanisms by which such restoration takes place. Thermodynamics dictates that if the enzyme-catalyzed reaction by which the modification was introduced is thermodynamically favorable, the free energy change involved in simply trying to run the reaction in reverse will be unfavorable. The phosphorylation of proteins on seryl, threonyl, or tyrosyl residues, catalyzed by protein kinases, is thermodynamically favored as a consequence of utilizing the high-energy gamma phosphoryl group of ATP. Phosphate groups are removed, not by recombining the phosphate with ADP to form ATP, but by a hydrolytic reaction catalyzed by enzymes called protein phosphatases. Similarly, acetyltransferases employ a high-energy donor substrate, NAD<sup>+</sup>, while deacetylases catalyze a direct hydrolysis that generates free acetate.

Because the high entropic barrier prevents the reunification of the two portions of a protein produced by hydrolysis of a peptide bond, proteolysis constitutes a physiologically irreversible modification. Once a proprotein is activated, it will continue to carry out its catalytic or other functions until it is removed by degradation or some other means. Zymogen activation thus represents a simple and economical, albeit one way, mechanism for restraining the latent activity of a protein until the appropriate circumstances are encountered. It is therefore not surprising that partial proteolysis is employed frequently to regulate proteins that work in the gastrointestinal tract or bloodstream rather than in the interior of cells.

## PROTEASES MAY BE SECRETED AS CATALYTICALLY INACTIVE PROENZYMES

Certain proteins are synthesized and secreted as inactive precursor proteins known as **proteins**. Selective, or "partial," proteolysis converts a proprotein by one or more successive proteolytic "clips" to a form that exhibits the characteristic activity of the mature protein, for example, its catalytic activity. The proprotein forms of enzymes are termed **proenzymes** or **zymogens**. Proteins synthesized as proenzymes include the hormone insulin (proprotein = proinsulin), the digestive enzymes pepsin, trypsin, and chymotrypsin (proproteins = pepsinogen, trypsinogen, and chymotrypsinogen, respectively), several factors of the blood clotting and blood clot dissolution cascades (see Chapter 51), and the connective tissue protein collagen (proprotein = procollagen).

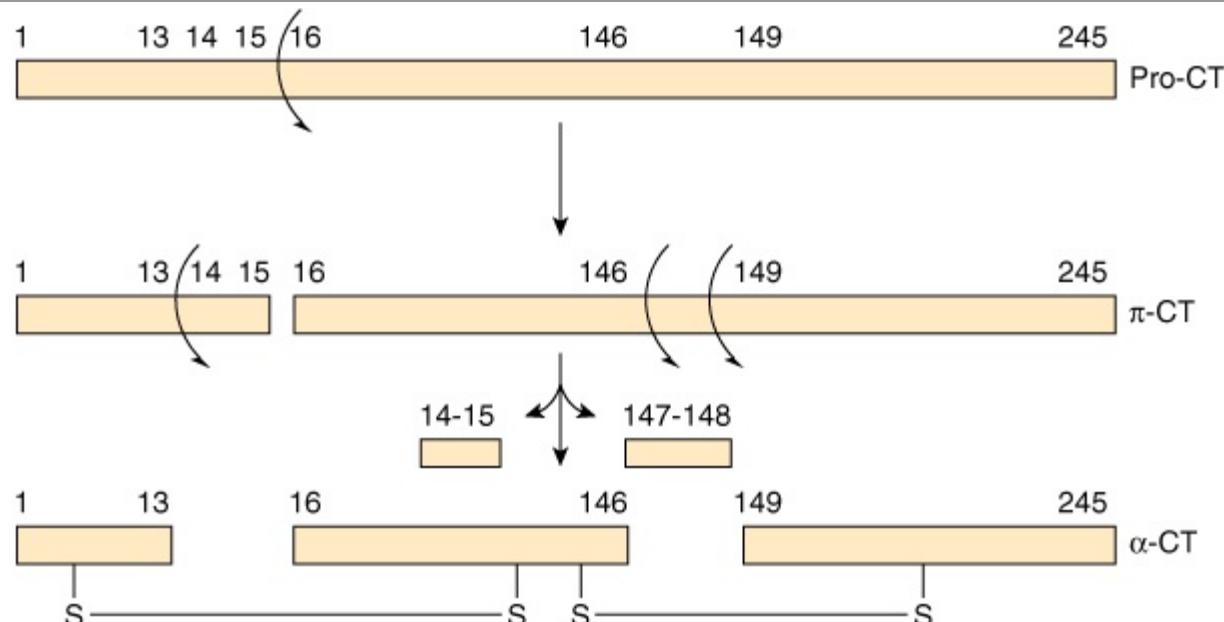
### Proenzymes Facilitate Rapid Mobilization of an Activity in Response to Physiologic Demand

The synthesis and secretion of proteases as catalytically inactive proenzymes protect the tissue of origin (eg, the pancreas) from autodigestion, such as can occur in pancreatitis. Certain physiologic processes such as digestion are intermittent but fairly regular and predictable in frequency. Others such as blood clot formation, clot dissolution, and tissue repair are brought "on line" only in response to pressing physiologic or pathophysiologic need. The processes of blood clot formation and dissolution clearly must be temporally coordinated to achieve homeostasis. Enzymes needed intermittently but rapidly often are secreted in an initially inactive form since new synthesis and secretion of the required proteins might be insufficiently rapid to respond to a pressing pathophysiologic demand such as the loss of blood (see Chapter 51).

### Activation of Prochymotrypsin Requires Selective Proteolysis

Selective proteolysis involves one or more highly specific proteolytic clips that may or may not be accompanied by separation of the resulting peptides. Most importantly, selective proteolysis often results in conformational changes that "create" the catalytic site of an enzyme. Note that while the catalytically essential residues His 57 and Asp 102 reside on the B peptide of  $\pi$ -chymotrypsin, Ser 195 resides on the C peptide (Figure 9-6). The conformational changes that accompany selective proteolysis of prochymotrypsin (chymotrypsinogen) align the three residues of the charge-relay network (see Figure 7-7), forming the catalytic site. Note also that contact and catalytic residues can be located on different peptide chains but still be within bond-forming distance of bound substrate.

**Figure 9-6**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

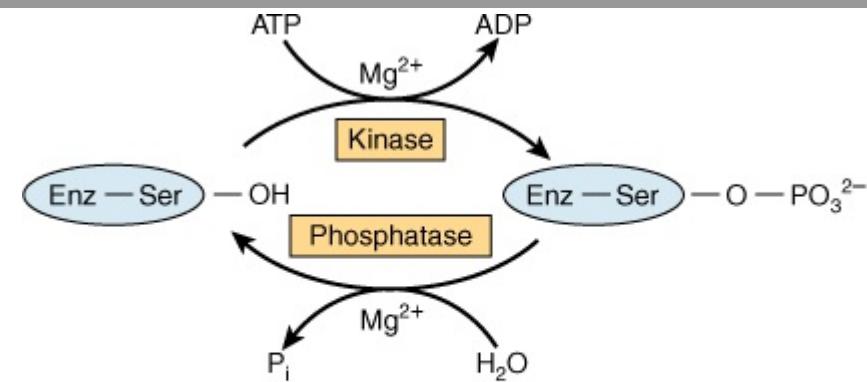
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two-dimensional representation of the sequence of proteolytic events that ultimately result in formation of the catalytic site of chymotrypsin, which includes the Asp 102-His57-Ser195 catalytic triad (see Figure 7-7). Successive proteolysis forms prochymotrypsin (pro-CT),  $\pi$ -chymotrypsin ( $\pi$ -CT), and ultimately  $\alpha$ -chymotrypsin ( $\alpha$ -CT), an active protease whose three peptides (A, B, C) remain associated by covalent inter-chain disulfide bonds.

## REVERSIBLE COVALENT MODIFICATION REGULATES KEY MAMMALIAN PROTEINS

Mammalian proteins are the targets of a wide range of covalent modification processes. Modifications such as prenylation, glycosylation, hydroxylation, and fatty acid acylation introduce unique structural features into newly synthesized proteins that tend to persist for the lifetime of the protein. Among the covalent modifications that regulate protein function (eg, methylation, acetylation), the most common by far is phosphorylation-dephosphorylation. **Protein kinases** phosphorylate proteins by catalyzing transfer of the terminal phosphoryl group of ATP to the hydroxyl groups of seryl, threonyl, or tyrosyl residues, forming O-phosphoseryl, O-phosphothreonyl, or O-phosphotyrosyl residues, respectively (**Figure 9–7**). Some protein kinases target the side chains of histidyl, lysyl, arginyl, and aspartyl residues. The unmodified form of the protein can be regenerated by hydrolytic removal of phosphoryl groups, catalyzed by **protein phosphatases**.

**Figure 9–7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Covalent modification of a regulated enzyme by phosphorylation-dephosphorylation of a seryl residue.**

A typical mammalian cell possesses thousands of phosphorylated proteins and several hundred protein kinases and protein phosphatases that catalyze their interconversion. The ease of interconversion of enzymes between their phospho- and dephospho- forms accounts, in part, for the frequency with which phosphorylation-dephosphorylation is utilized as a mechanism for regulatory control. Phosphorylation-dephosphorylation permits the functional properties of the affected enzyme to be altered only for as long as it serves a specific need. Once the need has passed, the enzyme can be converted back to its original form, poised to respond to the next stimulatory event. A second factor underlying the widespread use of protein phosphorylation-dephosphorylation lies in the chemical properties of the phosphoryl group itself. In order to alter an enzyme's functional properties, any modification of its chemical structure must influence the protein's three-dimensional configuration. The high charge density of protein-bound phosphoryl groups—generally -2 at physiologic pH—and their propensity to form strong salt bridges with arginyl and lysyl residues renders them potent agents for modifying protein structure and function. Phosphorylation generally influences an enzyme's intrinsic catalytic efficiency or other properties by inducing conformational changes. Consequently, the amino acids targeted by phosphorylation can be and typically are relatively distant from the catalytic site itself.

### Covalent Modifications Regulate Metabolite Flow

In many respects, sites of protein phosphorylation and other covalent modifications can be considered another form of allosteric site. However, in this case, the "allosteric ligand" binds covalently to the protein. Both phosphorylation-dephosphorylation and feedback inhibition provide short-term, readily reversible regulation of metabolite flow in response to specific physiologic signals. Both act without altering gene expression. Both act on early enzymes of a protracted biosynthetic metabolic pathway, and both act at allosteric rather than catalytic sites. Feedback inhibition, however, involves a single protein and lacks hormonal and neural features. By contrast, regulation of mammalian enzymes by phosphorylation-dephosphorylation involves several proteins and ATP, and is under direct neural and hormonal control.

## PROTEIN PHOSPHORYLATION IS EXTREMELY VERSATILE

Protein phosphorylation–dephosphorylation is a highly versatile and selective process. Not all proteins are subject to phosphorylation, and of the many hydroxyl groups on a protein's surface, only one or a small subset are targeted. While the most common enzyme function affected is the protein's catalytic efficiency, phosphorylation can also alter its location within the cell, susceptibility to proteolytic degradation, or responsiveness to regulation by allosteric ligands. Phosphorylation can increase an enzyme's catalytic efficiency, converting it to its active form in one protein, while phosphorylation of another protein converts it to an intrinsically inefficient, or inactive, form (**Table 9–1**).

**Table 9–1 Examples of Mammalian Enzymes Whose Catalytic Activity Is Altered by Covalent Phosphorylation–Dephosphorylation**

Enzyme	Activity State	
	Low	High
Acetyl-CoA carboxylase	EP	E
Glycogen synthase	EP	E
Pyruvate dehydrogenase	EP	E
HMG-CoA reductase	EP	E
Glycogen phosphorylase	E	EP
Citrate lyase	E	EP
Phosphorylase b kinase	E	EP
HMG-CoA reductase kinase	E	EP

**Abbreviations:** E, dephosphoenzyme; EP, phosphoenzyme.

Many proteins can be phosphorylated at multiple sites. Others are subject to regulation both by phosphorylation–dephosphorylation and by the binding of allosteric ligands, or by phosphorylation–dephosphorylation and another covalent modification. Phosphorylation–dephosphorylation at any one site can be catalyzed by multiple protein kinases or protein phosphatases. Many protein kinases and most protein phosphatases act on more than one protein and are themselves interconverted between active and inactive forms by the binding of second messengers or by covalent modification by phosphorylation–dephosphorylation.

The interplay between protein kinases and protein phosphatases, between the functional consequences of phosphorylation at different sites, between phosphorylation sites and allosteric sites, or between phosphorylation sites and other sites of covalent modification provides the basis for regulatory networks that integrate multiple environmental input signals to evoke an appropriate coordinated cellular response. In these sophisticated regulatory networks, individual enzymes respond to different environmental signals. For example, if an enzyme can be phosphorylated at a single site by more than one protein kinase, it can be converted from a catalytically efficient to an inefficient (inactive) form, or vice versa, in response to any one of several signals. If the protein kinase is activated in response to a signal different from the signal that activates the protein phosphatase, the phosphoprotein becomes a decision node. The functional output, generally catalytic activity, reflects the phosphorylation state. This state or degree of phosphorylation is determined by the relative activities of the protein kinase and protein phosphatase, a reflection of the presence and relative strength of the environmental signals that act through each.

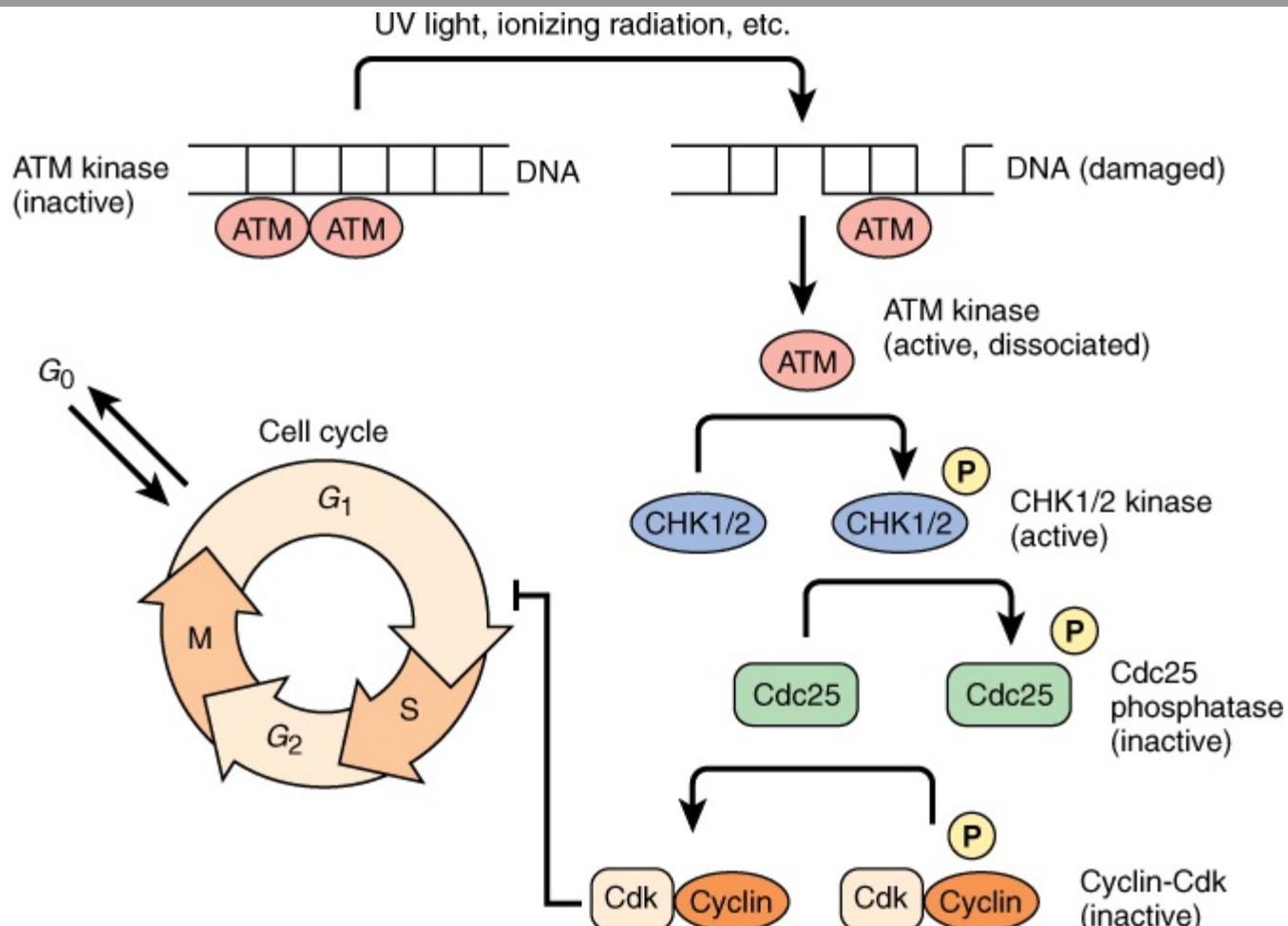
The ability of many protein kinases and protein phosphatases to target more than one protein provides a means for an environmental signal to coordinately regulate multiple metabolic processes. For example, the enzymes 3-hydroxy-3-methylglutaryl-CoA reductase and acetyl-CoA carboxylase—the rate-controlling enzymes for cholesterol and fatty acid biosynthesis, respectively—are phosphorylated and inactivated by the AMP-activated protein kinase. When this protein kinase is activated either through phosphorylation by yet another protein kinase or in response to the binding of its allosteric activator 5'-AMP, the two major pathways responsible for the synthesis of lipids from acetyl-CoA are both inhibited.

## INDIVIDUAL REGULATORY EVENTS COMBINE TO FORM SOPHISTICATED CONTROL NETWORKS

Cells carry out a complex array of metabolic processes that must be regulated in response to a broad spectrum of environmental factors. Hence, interconvertible enzymes and the enzymes responsible for their interconversion do not act as isolated "on" and "off" switches. In order to meet the demands of maintaining homeostasis, these building blocks are linked to form integrated regulatory networks.

One well-studied example of such a network is the eukaryotic cell cycle that controls cell division. Upon emergence from the G<sub>0</sub> or quiescent state, the extremely complex process of cell division proceeds through a series of specific phases designated G<sub>1</sub>, S, G<sub>2</sub>, and M (Figure 9–8). Elaborate monitoring systems, called checkpoints, assess key indicators of progress to ensure that no phase of the cycle is initiated until the prior phase is complete. Figure 9–8 outlines, in simplified form, part of the checkpoint that controls the initiation of DNA replication, called the S phase. A protein kinase called ATM is associated with the genome. If the DNA contains a double-stranded break, the resulting change in the conformation of the chromatin activates ATM. Upon activation, one subunit of the activated ATM dimer dissociates and initiates a series, or cascade, of protein phosphorylation–dephosphorylation events mediated by the CHK1 and CHK2 protein kinases, the Cdc25 protein phosphatase, and finally a complex between a cyclin and a cyclin-dependent protein kinase, or Cdk. Activation of the Cdk–cyclin complex blocks the G<sub>1</sub> to S transition, thus preventing the replication of damaged DNA. Failure at this checkpoint can lead to mutations in DNA that may lead to cancer or other diseases. Each step in the cascade provides a conduit for monitoring additional indicators of cell status prior to entering S phase.

**Figure 9–8**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A simplified representation of the G<sub>1</sub> to S checkpoint of the eukaryotic cell cycle. The circle shows the various stages in the eukaryotic cell cycle. The genome is replicated during S phase, while the two copies of the genome are segregated and cell division occurs during M phase. Each of these phases is separated by a G, or growth, phase characterized by an increase in cell size and the accumulation of the precursors required for the assembly of the large macromolecular complexes formed during S and M phases.

## SUMMARY

- Homeostasis involves maintaining a relatively constant intracellular and intra-organ environment despite wide fluctuations in the external environment. This is achieved via appropriate changes in the rates of biochemical reactions in response to physiologic need.
- The substrates for most enzymes are usually present at a concentration close to their  $K_m$ . This facilitates passive control of the rates of product formation in response to changes in levels of metabolic intermediates.
- Active control of metabolite flux involves changes in the concentration, catalytic activity, or both of an enzyme that catalyzes a committed, rate-limiting reaction.
- Selective proteolysis of catalytically inactive proenzymes initiates conformational changes that form the active site. Secretion as an inactive proenzyme facilitates rapid mobilization of activity in response to injury or physiologic need and may protect the tissue of origin (eg, autodigestion by proteases).
- Binding of metabolites and second messengers to sites distinct from the catalytic site of enzymes triggers conformational changes that alter  $V_{max}$  or  $K_m$ .
- Phosphorylation by protein kinases of specific seryl, threonyl, or tyrosyl residues—and subsequent dephosphorylation by protein phosphatases—regulates the activity of many human enzymes. The protein kinases and phosphatases that participate in regulatory cascades that respond to hormonal or second messenger signals constitute regulatory networks that can process and integrate complex environmental information to produce an appropriate and comprehensive cellular response.

## REFERENCES

- Ciechanover A, Schwartz AL: The ubiquitin system: pathogenesis of human diseases and drug targeting. *Biochim Biophys Acta* 2004;1695:3.
- Elgin SC, Reuter G. In Allis CD, Jenuwein T, Reinberg D, et al (editors): *Epigenetics*, Cold Spring Harbor Laboratory Press, 2007.
- Johnson LN, Lewis RJ: Structural basis for control by phosphorylation. *Chem Rev* 2001;101:2209.
- Muoio DM, Newgard CB: Obesity-related derangements in metabolic regulation. *Annu Rev Biochem* 2006;75:403.
- Scriven CR, Sly WS, Childs B, et al (editors): *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed. McGraw-Hill, 2000.
- Stiegitz K, Stec B, Baker DP, et al: Monitoring the transition from the T to the R state in *E. coli* aspartate transcarbamoylase by x-ray crystallography: crystal structures of the E50A mutant enzyme in four distinct allosteric states. *J Mol Biol* 2004; 341:853.
- Tu BP, Kudlicki A, Rowicka M, et al: Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* 2005;310:1152.
- Walsh CT: *Posttranslational Modification of Proteins. Expanding Nature's Inventory*, Roberts and Company Publishers, 2006.

## OBJECTIVES

After studying this chapter, you should be able to:

- Describe the major features of genomics, proteomics, and bioinformatics.
- Summarize the principal features and medical relevance of the Encode project.
- Describe the functions served by HapMap, *Entrez Gene*, BLAST, and the dbGAP databases.
- Describe the major features of computer-aided drug design and discovery.
- Describe possible future applications of computational models of individual pathways and pathway networks.
- Outline the possible medical utility of "virtual cells."

## BIOMEDICAL IMPORTANCE

The first scientific models of pathogenesis, such as Louis Pasteur's seminal germ theory of disease, were binary in nature: each disease possessed a single, definable causal agent. Malaria was caused by the amoeba *Plasmodium falciparum*, tuberculosis by the bacterium *Mycobacterium tuberculosis*, sickle cell disease by a mutation in a gene encoding one of the subunits of hemoglobin, poliomyelitis by poliovirus, and scurvy by a deficiency in ascorbic acid. The strategy for treating or preventing disease thus could be reduced to a straightforward process of tracing the causal agent, and then devising some means of eliminating it, neutralizing its effects, or blocking its route of transmission. This approach has been successfully employed to understand and treat a wide range of infectious and genetic diseases. However, it has become clear that the determinants of many pathologies—including cancer, coronary heart disease, type II diabetes, and Alzheimer's disease—are **multifactorial** in nature. Rather than having a specific causal agent or agents whose presence is both necessary and sufficient, the appearance and progression of the aforementioned diseases reflect the complex interplay between each individual's genetic makeup, other inherited or **epigenetic** factors, and environmental factors such as diet, lifestyle, toxins, viruses, or bacteria.

The challenge posed by **multifactorial diseases** demands a quantum increase in the breadth and depth of our knowledge of living organisms capable of matching their sophistication and complexity. We must identify the many as yet unknown proteins encoded within the genomes of humans and the organisms with which they interact, their cellular functions and interactions. We must be able to trace the factors, both external and internal, that compromise human health and wellbeing by analyzing the impact of dietary, genetic, and environmental factors across entire communities or populations. The sheer mass of information that must be processed lies well beyond the ability of the human mind to review and analyze unaided. To understand, as completely and comprehensively as possible, the molecular mechanisms that underlie the behavior of living organisms, the manner in which perturbations can lead to disease or dysfunction, and how such perturbing factors spread throughout a population, biomedical scientists have turned to sophisticated computational tools to collect and evaluate biologic information on a mass scale.

## GENOMICS: AN INFORMATION AVALANCHE

Physicians and scientists have long understood that the genome, the complete complement of genetic information of a living organism, represented a rich source of information concerning topics ranging from basic metabolism to evolution to aging. However, the massive size of the human genome,  $3 \times 10^9$  nucleotide base pairs, required a paradigm shift in the manner in which scientists approached the determination of DNA sequences. Recent advances in bioinformatics and computational biology have, in turn, been fueled by the need to develop new approaches to "mine" the mass of sequence data generated by the application of increasingly efficient and economical technology to the genomes of hundreds of new organisms and, most recently, of several individual human beings.

### The Human Genome Project

The successful completion of the Human Genome Project (HGP) represents the culmination of more than six decades of achievements in molecular biology, genetics, and biochemistry. The chronology below lists several of the milestone events that led to the determination of the entire sequence of the human genome.

- 1944—DNA is shown to be the hereditary material
- 1953—Concept of the double helix is posited
- 1966—The genetic code is solved
- 1972—Recombinant DNA technology is developed
- 1977—Practical DNA sequencing technology emerges
- 1983—The gene for Huntington's disease is mapped
- 1985—The Polymerase Chain Reaction (PCR) is invented
- 1986—DNA sequencing becomes automated
- 1986—The gene for Duchenne muscular dystrophy is identified
- 1989—The gene for cystic fibrosis is identified
- 1990—The Human Genome Project is launched in the United States
- 1994—Human genetic mapping is completed
- 1996—The first human gene map is established
- 1999—The Single Nucleotide Polymorphism Initiative is started
- 1999—The first sequence of a human chromosome, number 22, is completed
- 2000—"First draft" of the human genome is completed
- 2003—Sequencing of the first human genome is completed
- 2010—Scientists embark on the sequencing of 1000 individual genomes to determine degree of genetic diversity in humans

By 2011 >180 eukaryotic, prokaryotic, and archaeal genomes had been sequenced. Examples include the genomes of chicken, cat, dog, elephant, rat, rabbit, orangutan, woolly mammoth, and duck-billed platypus, and the genomes of several individuals including Craig Venter and James Watson. The year 2010 saw completion of the Neanderthal genome, whose initial analysis suggests that up to 2% of the DNA in the genome of present-day humans outside of Africa originated in Neanderthals or in Neanderthal ancestors.

As of this writing, the genome sequences of >5,000 biological entities, ranging from viruses and bacteria to plants and animals have been reported. Ready access to genome sequences from organisms spanning all three phylogenetic domains and to the powerful algorithms requisite for manipulating and transforming the data derived from these sequences has transformed basic research in biology, microbiology, pharmacology, and biochemistry.

### Genomes and Medicine

By comparing the genomes of pathogenic and nonpathogenic strains of a particular microorganism, genes likely to encode determinants of virulence can be highlighted by virtue of their presence in only the virulent strain. Similarly, comparison of the genomes of a pathogen with its host can identify genes unique to the former. Drugs targeting the protein products of the pathogen-specific genes should, in theory, produce little or no side effects for the infected host. The coming decade will witness the expansion of the "Genomics Revolution" into the day-to-day practice of medicine and agriculture as physicians and scientists exploit new knowledge of the human genome and of the genomes of the organisms that colonize, feed, and infect *Homo sapiens*. Whereas the first human genome project required several years, hundreds of people, and many millions of dollars to complete, quantum leaps in efficiency and economy have led one company to project that up to one million persons will have their individual genome sequences determined by the year 2014. The ability to diagnose and treat patients guided by knowledge of their genetic makeup, an approach popularly referred to as "designer medicine," will render medicine safer and more effective.

### Exome Sequencing

A harbinger of this new era has been provided by the application of "exome sequencing" to the diagnosis of rare or cryptic genetic diseases. The exome consists of those segments of DNA, called exons, that code for the amino acid sequences of proteins (Chapter 36). Since exons comprise only about 1% of the human genome, the exome represents a much smaller and more tractable target than the complete genome. Comparison of exome sequences has identified the genes responsible for a growing list of diseases that includes retinitis pigmentosa, Freeman-Sheldon syndrome, and Kabuki syndrome.

### The Potential Challenges of Designer Medicine

While genome-based "designer medicine" promises to be very effective, it also confronts humanity with profound challenges in the areas of ethics, law, and public policy. Who owns and controls access to this information? Can a life or health insurance company, for instance, deny coverage to an individual based upon the risk factors apparent from their genome sequence? Does a prospective employer have the right to know a potential employee's genetic makeup? Do prospective spouses have the right to know their fiancées' genetic risk factors? Ironically, the resolution of these issues may prove a more lengthy and laborious process than did the determination of the first human genome sequence.

## BIOINFORMATICS

**Bioinformatics** exploits the formidable information storage and processing capabilities of the computer to develop tools for the collection, collation, retrieval, and analysis of biologic data on a mass scale. That many bioinformatic resources (see below) can be accessed via the Internet provides them with global reach and impact. The central objective of a typical bioinformatics project is to assemble all of the available information relevant to a particular topic in a single location, often referred to as a **library** or **database**, in a uniform format that renders the data amenable to manipulation and analysis by computer algorithms.

### Bioinformatic Databases

The size and capabilities of bioinformatic databases can vary widely depending upon the scope and nature of their objectives. The PubMed database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>) compiles citations for all articles published in thousands of journals devoted to biomedical and biological research. Currently, PubMed contains >20 million citations. By contrast, the RNA Helicase Database (<http://www.mahelicase.org/>) confines itself to the sequence, structure, and biochemical and cellular function of a single family of enzymes, the RNA helicases.

### Challenges of Database Construction

The construction of a comprehensive and user-friendly database presents many challenges. First, biomedical information comes in a wide variety of forms. For example, the coding information in a genome, although voluminous, is composed of simple linear sequences of four nucleotide bases. While the number of amino acid residues that define a protein's primary structure is minute relative to the number of base pairs in a genome, a description of a protein's x-ray structure requires that the location of each atom be specified in three-dimensional space.

Second, the designer must correctly anticipate the manner in which users may wish to search or analyze the information within a database, and devise algorithms for coping with these variables. Even the seemingly simple task of searching a gene database commonly employs, alone or in various combinations, criteria as diverse as the name of the gene, the name of the protein that it encodes, the biologic function of the gene product, a nucleotide sequence within the gene, a sequence of amino acids within the protein it encodes, the organism in which it is present, or the name of the investigators who determined the sequence of that gene.

## EPIDEMIOLOGY ESTABLISHED THE MEDICAL POTENTIAL OF INFORMATION PROCESSING

The power of basic biomedical research resides in the laboratory scientist's ability to manipulate homogenous, well-defined research targets under carefully controlled circumstances. The ability to independently vary the qualitative and quantitative characteristics of both target and input variables permits cause-effect relationships to be determined in a direct and reliable manner. These advantages are obtained, however, by employing "model" organisms such as mice or cultured human cell lines as stand-ins for the human patients that represent the ultimate targets for, and beneficiaries of, this research. Laboratory animals do not always react as do *Homo sapiens*, nor can a dish of cultured fibroblast, kidney, or other cells replicate the incredible complexity of a human being.

Although unable to conduct rigorously controlled experiments on human subjects, careful observation of real world behavior has long proven to be a source of important biomedical insights. Hippocrates, for example, noted that while certain **epidemic** diseases appeared in a sporadic fashion, **endemic** diseases such as malaria exhibited clear association with particular locations, age group, etc. **Epidemiology** refers to the branch of the biomedical sciences that employs bioinformatic approaches to extend our ability and increase the accuracy with which we can identify factors that contribute to or detract from human health through the study of real world populations.

### Early Epidemiology of Cholera

One of the first recorded epidemiological studies, conducted by Dr. John Snow, employed simple geospatial analysis to track the source of a cholera outbreak. Epidemics of cholera, typhus, and other infectious diseases were relatively common in the crowded, unsanitary conditions of nineteenth century London. By mapping the locations of the victims' residences, Snow was able to trace the source of the contagion to the contamination of one of the public pumps that supplied citizens with their drinking water (Figure 10-1). Unfortunately, the limited capacity of hand calculations or graphing rendered the success of analyses such as Snow's critically dependent upon the choice of the working hypothesis used to select the variables to be measured and processed. Thus, while nineteenth century Londoners also widely recognized that haberdashers were particularly prone to display erratic and irrational behavior (eg, "as Mad as a Hatter"), nearly a century would pass before the cause was traced to the mercury compounds used to prepare the felt from which the hats were constructed.

**Figure 10-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: Harper's Illustrated Biochemistry, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

This version of the map drawn by Dr. John Snow compares the location of the residences of victims of an 1854 London cholera epidemic (Dots) with the

**locations of the pumps that supplied their drinking water (X's).** Contaminated water from the pump on Broad Street, lying roughly in the center of the cluster of victims, proved to be the source of the epidemic in this neighborhood.

## Impact of Bioinformatics on Epidemiological Analysis

As the process of data analysis has become automated, the sophistication and success rate of epidemiological analyses has risen accordingly. Today, complex computer algorithms enable researchers to assess the influence of a broad range of health-related parameters when tracking the identity and source or reconstructing the transmission of a disease or condition: height; weight; age; gender; body mass index; diet; ethnicity; medical history; profession; drug, alcohol, or tobacco use; exercise; blood pressure; habitat; marital status; blood type; serum cholesterol level; etc. Equally important, modern bioinformatics may soon enable epidemiologists to dissect the identities and interactions of the multiple factors underlying complex diseases such as cancer or Alzheimer's disease.

The accumulation of individual genome sequences will shortly introduce a powerful new dimension to the host of biological, environmental, and behavioral factors to be compared and contrasted with each person's medical history. One of the first fruits of these studies has been the identification of genes responsible for a few of the >3000 known or suspected Mendelian disorders whose causal genetic abnormalities have yet to be traced. The ability to evaluate contributions of and the interactions among an individual's genetic makeup, behavior, environment, diet, and lifestyle holds the promise of eventually revealing the answers to the age-old question of why some persons exhibit greater vitality, stamina, longevity, and resistance to disease than others—in other words, the root sources of health and wellness.

## BIOINFORMATIC AND GENOMIC RESOURCES

The large collection of databases that have been developed for the assembly, annotation, analysis and distribution of biological and biomedical data reflects the breadth and variety of contemporary molecular, biochemical, epidemiological, and clinical research. Below are listed examples of the following prominent bioinformatics resources: UniProt, GenBank, and the Protein Database (PDB) represent three of the oldest and most widely used bioinformatics databases. Each complements the other by focusing on a different aspect of macromolecular structure.

### Uniprot

The UniProt Knowledgebase (<http://www.pir.uniprot.org/>) can trace its origins to the *Atlas of Protein Sequence and Structure*, a printed encyclopedia of protein sequences first published in 1968 by Margaret Dayhoff and the National Biomedical Research Foundation at Georgetown University. The aim of the Atlas was to facilitate studies of protein evolution using the amino acid sequences being generated consequent to the development of the Edman method for protein sequencing (Chapter 4). In partnership with the Martinsreid Institute for Protein Sequences and the International Protein Information Database of Japan, the Atlas made the transition to electronic form as the Protein Information Resource (PIR) Protein Sequence Database in 1984. In 2002, PIR integrated its database of protein sequence and function with the Swiss-Prot Protein Database established by Amos Bairoch in 1986 under the auspices of the Swiss Institute of Bioinformatics and the European Bioinformatics Institute, to form the world's most comprehensive resource on protein structure and function, the UniProt Knowledgebase.

### Genbank

The goal of GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>), the National Institutes of Health's (NIH) genetic sequence database, is to collect and store all known biological nucleotide sequences and their translations in a searchable form. Established in 1979 by Walter Goad of Los Alamos National Laboratory, GenBank currently is maintained by the National Center for Biotechnology Information at the NIH. GenBank constitutes one of the cornerstones of the International Sequence Database Collaboration, a consortium that includes the DNA Database of Japan and the European Molecular Biology Laboratory.

### PDB

The RCSB Protein Data Base (PDB) (<http://www.rcsb.org/pdb/home/home.do>), a repository of the three-dimensional structures of proteins, polynucleotides, and other biological macromolecules, was established in 1971 by Edgar Meyer and Walter Hamilton of Brookhaven National Laboratories. In 1998, responsibility for the PDB was transferred to the Research Collaboration for Structural Bioinformatics formed by Rutgers University, the University of California at San Diego, and the University of Wisconsin. The PDB contains well over 50,000 three-dimensional structures for proteins, as well as proteins bound with substrates, substrate analogs, inhibitors, or other proteins. The user can rotate these structures freely in three-dimensional space, highlight specific amino acids, and select from a variety of formats such as space filling, ribbon, backbone, etc. (Chapters 5, 6, and 10).

### SNPs & Tagged SNPs

While the genome sequence of any two individuals is 99.9% identical, human DNA contains ~10 million sites where individuals differ by a single-nucleotide base. These sites are called **Single-Nucleotide Polymorphisms** or **SNPs**. When sets of SNPs localized to the same chromosome are inherited together in blocks, the pattern of SNPs in each block is termed a **haplotype**. By comparing the haplotype distributions between groups of individuals that differ in some physiological characteristic, such as susceptibility to a disease, biomedical scientists can identify SNPs that are associated with specific phenotypic traits. This process can be facilitated by focusing on **Tag SNPs**, a subset of the SNPs in a given block sufficient to provide a unique marker for a given haplotype. Detailed study of each region should reveal variants in genes that contribute to a specific disease or response.

### HapMap

In 2002, scientists from the United States, Canada, China, Japan, Nigeria, and the United Kingdom launched the International **Haplotype Map (HapMap)** Project (<http://hapmap.ncbi.nlm.nih.gov/>), a comprehensive effort to identify SNPs associated with common human diseases and differential responses to pharmaceuticals. The resulting **HapMap Database** should lead to earlier and more accurate diagnosis, improved prevention, and patient management. Knowledge of an individual's genetic profile will also be used to guide the selection of safer and more effective drugs or vaccines, a process termed **pharmacogenomics**. These genetic markers will also provide labels with which to identify and track specific genes as scientists seek to learn more about the critical processes of genetic inheritance and selection.

## ENCODE

Identification of all the *functional elements* of the genome will vastly expand our understanding of the molecular events that underlie human development, health, and disease. To address this goal, in late 2003, the National Human Genome Research Institute (NHGRI) initiated the ENCODE (**Encyclopedia of DNA Elements**) Project. Based at the University of California at Santa Cruz, ENCODE (<http://www.genome.gov/10005107>) is a collaborative effort that combines laboratory and computational approaches to identify every functional element in the human genome. Consortium investigators with diverse backgrounds and expertise collaborate in the development and evaluation of new high-throughput techniques, technologies, and strategies to address current deficiencies in our ability to identify functional elements.

The pilot phase of ENCODE targeted ~1% (30 Mb) of the human genome for rigorous computational and experimental analysis. A variety of methods were employed to identify, or **annotate**, the function of each portion of the DNA in 500 base pair steps. These pilot studies revealed that the human genome contains a large number and variety of functionally active components interwoven to form complex networks. The successful prosecution of this pilot study has resulted in the funding of a series of Scale-Up Projects aimed at tackling the remaining 99% of the genome.

## Entrez Gene

**Entrez Gene** (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>), a database maintained by the National Center for Biotechnology Information (NCBI), provides a variety of information about individual human genes. The information includes the sequence of the genome in and around the gene, exon-intron boundaries, the sequence of the mRNA(s) produced from the gene, and any known phenotypes associated with a given mutation of the gene in question. Entrez Gene also lists, where known, the function of the encoded protein and the impact of known single-nucleotide polymorphisms within its coding region.

## dbGAP

**dbGAP**, the **Database of Genotype and Phenotype**, is an NCBI database that complements Entrez Gene (<http://www.ncbi.nlm.nih.gov/gap>). dbGAP compiles the results of research into the links between specific genotypes and phenotypes. To protect the confidentiality of sensitive clinical data, the information contained in dbGAP is organized into open- and controlled-access sections. Access to sensitive data requires that the user apply for authorization to a data access committee.

## Additional Databases

Other databases dealing with human genetics and health include OMIM, Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>), **HGMD**, the Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk/ac/index.php>), the Cancer Genome Atlas (<http://cancergenome.nih.gov/>), and GeneCards (<http://www.genecards.org/>), which tries to collect all relevant information on a given gene from databases worldwide to create a single, comprehensive "card" for each.

## COMPUTATIONAL BIOLOGY

The primary objective of **computational biology** is to develop computer models that apply physical, chemical, and biological principles to reproduce the behavior of biologic molecules and processes. Unlike bioinformatics, whose major focus is the collection and evaluation of existing data, computational biology is experimental and exploratory in nature. By performing virtual experiments and analyses "*in silico*," meaning performed on a computer or through a computer simulation, computational biology offers the promise of greatly accelerating the pace and efficiency of scientific discovery.

Computational biologists are attempting to develop predictive models that will (1) permit the three-dimensional structure of a protein to be determined directly from its primary sequence, (2) determine the function of unknown proteins from their sequence and structure, (3) screen for potential inhibitors of a protein *in silico*, and (4) construct virtual cells that reproduce the behavior and predict the responses of their living counterparts to pathogens, toxins, diet, and drugs. The creation of computer algorithms that accurately imitate the behavior of proteins, enzymes, cells, etc., promises to enhance the speed, efficiency, and the safety of biomedical research. Computational biology will also enable scientists to perform experiments *in silico* whose scope, hazard, or nature renders them inaccessible to, or inappropriate for, conventional laboratory or clinical approaches.

## IDENTIFICATION OF PROTEINS BY HOMOLOGY

One important method for the identification, also called **annotation**, of novel proteins and gene products compares the sequences of novel proteins with those of proteins whose functions or structures had been determined. Simply put, homology searches and multiple sequence comparisons operate on the principle that proteins that perform similar functions will share conserved domains or other sequence features or **motifs**, and vice versa (Figure 10–2). Of the many algorithms developed for this purpose, the most widely used is **BLAST**.

**Figure 10–2**

Language	Word	Alignment
English	PHYSIOLOGICAL	P H Y S I O L O G I - C A L
French	PHYSIOLOGIQUE	P H Y S I O L O G I - Q U E
German	PHYSIOLOGISCH	P H Y S I O L O G I S C H
Dutch	FYSIOLOGISCH	F - Y S I O L O G I S C H
Spanish	FYSIOLÓGICO	F - Y S I O L O G I - C O
Polish	FIZJOLOGICZNY	F - I Z J O L O G I - C Z N Y

Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Representation of a multiple sequence alignment.** Languages evolve in a fashion that mimics that of genes and proteins. Shown is the English word "physiological" in several languages. The alignment demonstrates their conserved features. Identities with the English word are shown in dark red; linguistic similarities in blue. Multiple sequence alignment algorithms identify conserved nucleotide and amino acid letters in DNA, RNA, and polypeptides in an analogous fashion.

## BLAST

**BLAST** (**B**asic **L**ocal **A**lignment **S**earch **T**ool) and other sequence comparison/alignment algorithms trace their origins to the efforts of early molecular biologists to determine whether observed similarities in sequence among proteins that perform parallel metabolic functions were indicative of progressive changes in a common ancestral protein. The major evolutionary question addressed was whether the similarities reflected (1) descent from a common ancestral protein (**divergent evolution**) or (2) the independent selection of a common mechanism for meeting some specific cellular need (**convergent evolution**), as would be anticipated if one particular solution was overwhelmingly superior to the alternatives. Calculation of the minimum number of nucleotide changes required to interconvert putative protein isoforms allows inferences to be drawn concerning whether or not the similarities and differences exhibit a pattern indicative of progressive change from a shared origin.

BLAST has evolved into a family of programs optimized to address specific needs and data sets. Thus, **blastp** compares an *amino acid* query sequence against a *protein* sequence database, **blastn** compares a *nucleotide* query sequence against a *nucleotide* sequence database, **blastx** compares a *nucleotide* query sequence translated in all reading frames against a *protein* sequence database to reveal potential translation products, **tblastn** compares a *protein* query sequence against a *nucleotide* sequence database dynamically translated in all six reading frames, and **tblastx** compares the six-frame translations of a *nucleotide* query sequence against the six-frame translations of a *nucleotide* sequence database. Unlike multiple sequence alignment programs that rely on *global* alignments, the **BLAST** algorithms emphasize regions of *local* alignment to detect relationships among sequences with only isolated regions of similarity. This approach provides speed and increased sensitivity for distant sequence relationships. Input or "query" sequences are broken into "words" (default size 11 for nucleotides, 3 for amino acids). Word hits to databases are then extended in both directions.

## IDENTIFICATION OF "UNKNOWN" PROTEINS

A substantial portion of the genes discovered by genome sequencing projects code for "unknown" or hypothetical polypeptides for which homologs of known function are lacking. Bioinformaticists are developing tools to enable scientists to deduce the three-dimensional structure and function of cryptic proteins directly from their amino acid sequences. Currently, the list of unknown proteins uncovered by genomics contains thousands of entries, with new entries being added as more genome sequences are solved. The ability to generate structures and infer function *in silico* promises to significantly accelerate protein identification and provide insight into the mechanism by which proteins fold. This knowledge will aid in understanding the underlying mechanisms of various protein folding diseases, and will assist molecular engineers to design new proteins to perform novel functions.

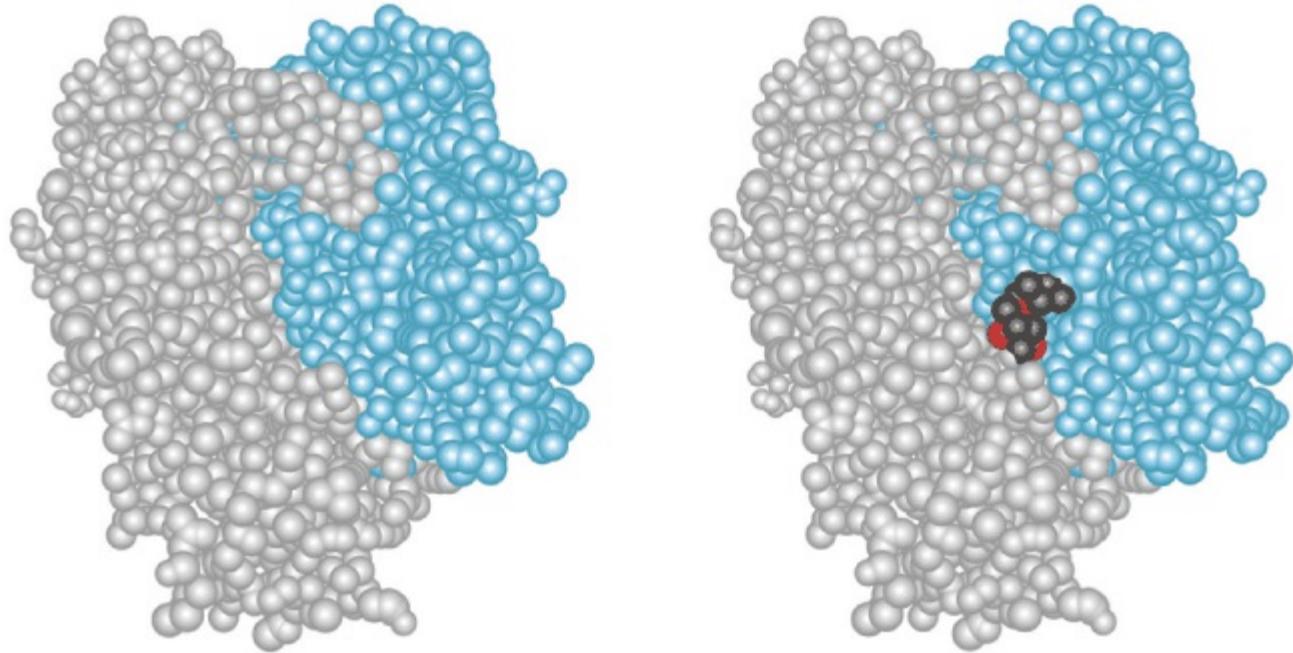
### The Folding Code

Comparison of proteins whose three-dimensional structures have been determined by x-ray crystallography or NMR spectroscopy can reveal patterns that link specific primary sequence features to specific primary, secondary, and tertiary structures—sometimes called the folding code. The first algorithms used the frequency with which individual amino acids occurred in  $\alpha$ -helices,  $\beta$ -sheets, turns, and loops to predict the number and location of these elements within the sequence of a polypeptide, known as secondary structure topography. By extending this process, for example, by weighing the impact of hydrophobic interactions in the formation of the protein core, algorithms of remarkable predictive reliability are being developed. However, while current programs perform well in generating the conformations of proteins comprised of a single domain, projecting the likely structure of membrane proteins and those composed of multiple domains remains problematic.

### Relating Three-Dimensional Structure to Function

Scientists are also attempting to discern patterns of three-dimensional structure that correlate to specific physiologic functions. The space-filling representation of the enzyme HMG-CoA reductase and its complex with the drug lovastatin (**Figure 10–3**) provides some perspective on the challenges inherent in identifying ligand-binding sites from scratch. Where a complete three-dimensional structure can be determined or predicted, the protein's surface can be scanned for the types of pockets and crevices indicative of likely binding sites for substrates, allosteric effectors, etc., by any one of a variety of methods such as tracing its surface with balls of a particular dimension (**Figure 10–4**). Surface maps generated with the program Graphical Representation and Analysis of Surface Properties, commonly referred to as **GRASP diagrams**, highlight the locations of neutral, negatively charged, and positively charged functional groups on a protein's surface (**Figure 10–5**) to infer a more detailed picture of the biomolecule that binds to or "docks" at that site. The predicted structure of the ligands that bind to an unknown protein, along with other structural characteristics and sequence motifs can then provide scientists with the clues needed to make an "educated guess" regarding its biological function(s).

**Figure 10–3**

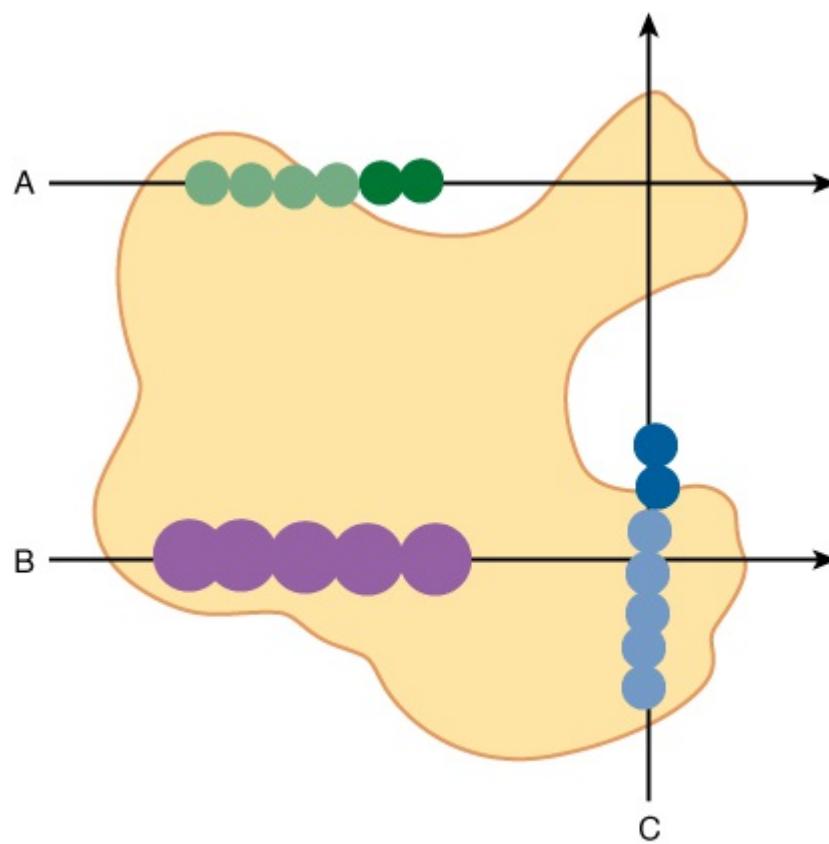


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Shown are space-filling representations of the homodimeric HMG-CoA reductase from *Pseudomonas mevalonii* with (right) and without (left) the statin drug lovastatin bound. Each atom is represented by a sphere the size of its van der Waals' radius. The two polypeptide chains are colored gray and blue. The carbon atoms of lovastatin are colored black and the oxygen atoms red. Compare this model with the backbone representations of proteins shown in Chapters 5 and 6. (Adapted from Protein Data Bank ID no. 1t02.)

**Figure 10–4**

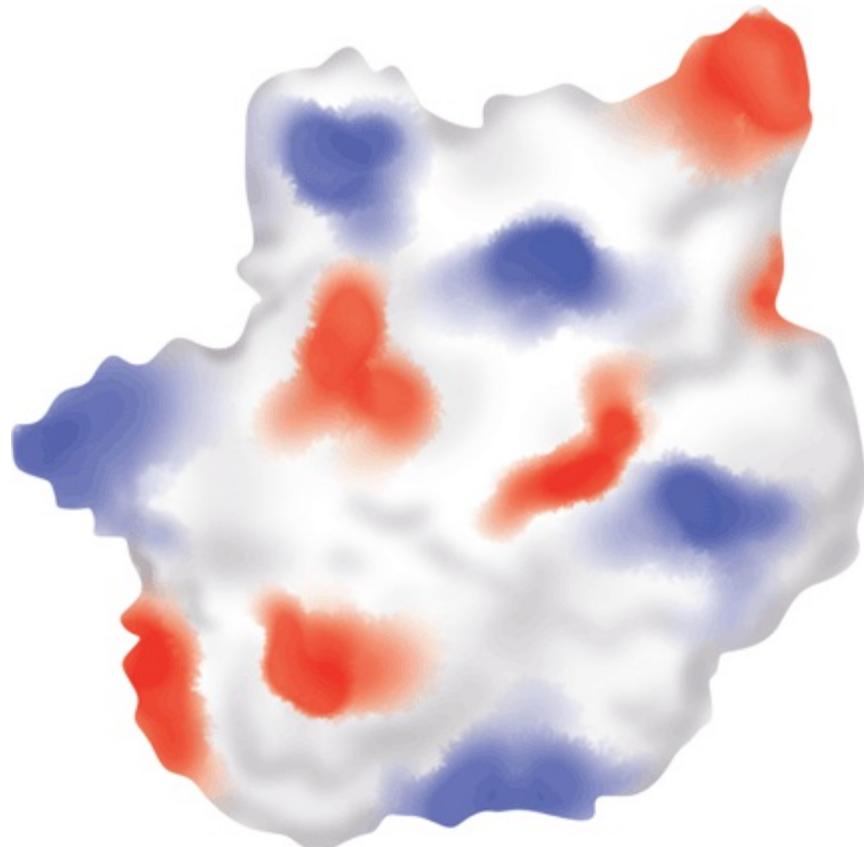


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**A simplified representation of a ligand site prediction program.** Ligand site prediction programs such as POCKET, LIGSITE, or Pocket-Finder convert the three-dimensional structure of a protein into a set of coordinates for its component atoms. A two-dimensional slice of the space filled by these coordinates is presented as an irregularly shaped outline (yellow). A round probe is then passed repeatedly through these coordinates along a series of lines paralleling each of the three coordinate axes (A, B, C). Lightly shaded circles represent positions of the probe where its radius overlaps with one or more atoms in the Cartesian coordinate set. Darkly shaded circles represent positions where no protein atom coordinates fall within the probe's radius. In order to qualify as a pocket or crevice within the protein, and not just open space outside of it, the probe must eventually encounter protein atoms lying on the other side of the opening (C).

**Figure 10–5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Representation of a GRASP diagram indicating the electrostatic topography of a protein.** Shown is a space-filling representation of a hypothetical protein. Areas shaded in

red indicate the presence of amino acid side chains or other moieties on the protein surface predicted to bear a negative charge at neutral pH. Blue indicates the presence of predicted positively charged groups. White denotes areas predicted to be electrostatically neutral.

## COMPUTER-AIDED DRUG DESIGN

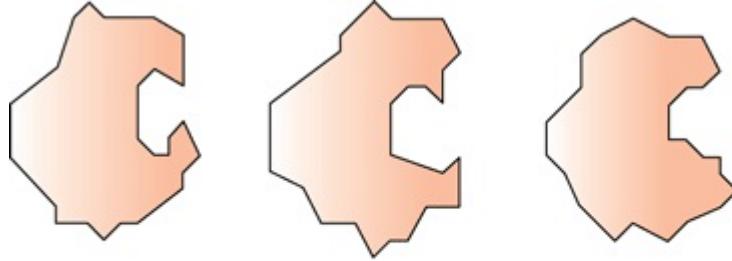
**Computer-Aided Drug Design (CADD)** employs the same type of molecular-docking algorithms used to identify ligands for unknown proteins. However, in this case the set of potential ligands to be considered is not confined to those occurring in nature and is aided by empirical knowledge of the structure or functional characteristics of the target protein.

### Molecular Docking Algorithms

For proteins of known three-dimensional structure, molecular-docking approaches employ programs that attempt to fit a series of potential ligand "pegs" into a designated binding site or "hole" on a protein template. To identify optimum ligands, docking programs must account for matching shapes as well as the presence and position of complementary hydrophobic, hydrophilic, and charge attributes. The binding affinities of the inhibitors selected on the basis of early docking studies were disappointing, as the rigid models for proteins and ligands employed were incapable of replicating the conformational changes that occur in both ligand and protein as a consequence of binding, a phenomenon referred to as "induced fit" (Chapter 7).

Imbuing proteins and ligands with conformational flexibility requires massive computing power, however. Hybrid approaches have thus evolved that employ a set, or ensemble, of templates representing slightly different conformations of the protein (**Figure 10–6**) and either ensembles of ligand conformers (**Figure 10–7**) or ligands in which only a few select bonds are permitted to rotate freely. Once the set of potential ligands has been narrowed, more sophisticated docking analyses can be undertaken to identify high-affinity ligands able to interact with their protein target across the latter's spectrum of conformational states.

**Figure 10–6**

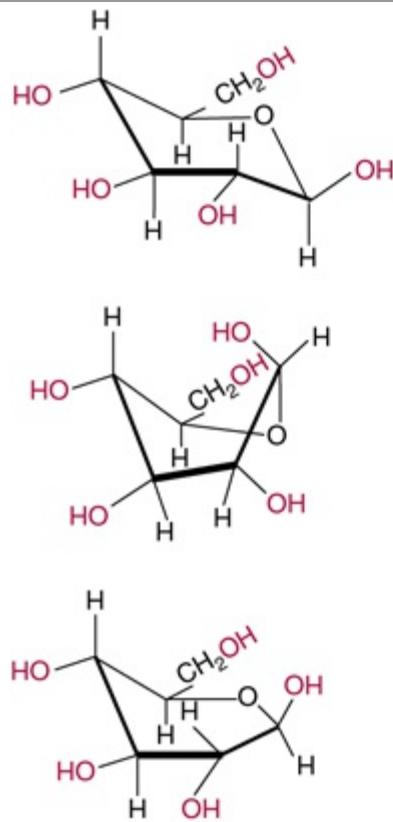


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two-dimensional representation of a set of conformers of a protein. Notice how the shape of the binding site changes.

**Figure 10–7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Conformers of a simple ligand.** Shown are three of the many different conformations of glucose, commonly referred to as chair (Top), twist boat (Middle), and half chair (Bottom). Note the differences not only in shape and compactness but in the position of the hydroxyl groups, potential participants in hydrogen bonds, as highlighted in red.

## Structure–Activity Relationships

If no structural template is available for the protein of interest, computers can be used to assist the search for high-affinity inhibitors by calculating and projecting **Structure–Activity Relationships (SARs)**. In this process, the measured binding affinities for several known inhibitors are compared and contrasted to determine whether specific chemical features make positive or negative thermodynamic contributions to ligand binding. This information can then be used to search databases of chemical compounds to identify those which possess the

most promising combination of positive versus negative features.

## SYSTEMS BIOLOGY & VIRTUAL CELLS

### The Goal of Systems Biology Is to Construct Molecular Circuit Diagrams

What if a scientist could detect, in a few moments, the effect of inhibiting a particular enzyme, or replacing a particular gene, the response of a muscle cell to insulin, the proliferation of a cancer cell, or the production of beta amyloid by entering the appropriate query into a computer? The goal of **systems biology** is to construct the molecular equivalent of circuit diagrams that faithfully depict the components of a particular functional unit and the interactions between them in logical or mathematical terms. These functional units can range in size and complexity from the enzymes and metabolites within a biosynthetic pathway to the network of proteins that controls the cell division cycle to, ultimately, entire cells, organs, and organisms. These models can then be used to perform "virtual" experiments that can enhance the speed and efficiency of empirical investigations by identifying the most promising lines of investigation and assisting in the evaluation of results. Perhaps more significantly, the ability to conduct virtual experiments extends the reach of the investigator, within the limits of the accuracy of the model, beyond the reach of current empirical technology.

Already, significant progress is being made. By constructing virtual molecular networks, scientists have been able to determine how cyanobacteria assemble a reliable circadian clock using only four proteins. Models of the T cell receptor signaling pathway have revealed how its molecular circuitry has been arranged to produce switch-like responses upon stimulation by agonist peptide-major histocompatibility complexes (MHC) on an antigen-presenting cell. Scientists can use the gaps encountered in modeling molecular and cellular systems to guide the identification and annotation of the remaining protein pieces, in the same way that someone who solves a jigsaw puzzle surveys the remaining pieces for matches to the gaps in the puzzle. This reverse engineering approach has been successfully used to define the function of type II glycerate 2-kinases in bacteria and to identify "cryptic" folate synthesis and transport genes in plants.

### Virtual Cells

Recently, scientists have been able to successfully create a sustainable metabolic network, composed of nearly two hundred proteins—an important step toward the creation of a **virtual cell**. The "holy grail" of systems biologists is to replicate the behavior of living human cells *in silico*. The potential benefits of such virtual cells are enormous. Not only will they permit optimum sites for therapeutic intervention to be identified in a rapid and unbiased manner, but unintended side effects may be revealed prior to the decision to invest time and resources in the synthesis, analysis, and trials of a potential pharmacophore. The ability to conduct fast, economic toxicological screening of materials ranging from herbicides to cosmetics will benefit human health. Virtual cells can also aid in diagnosis. By manipulating a virtual cell to reproduce the metabolic profile of a patient, underlying genetic abnormalities may be revealed. The interplay of the various environmental, dietary, and genetic factors that contribute to multifactorial diseases such as cancer can be systematically analyzed. Preliminary trials of potential gene therapies can be assessed safely and rapidly *in silico*.

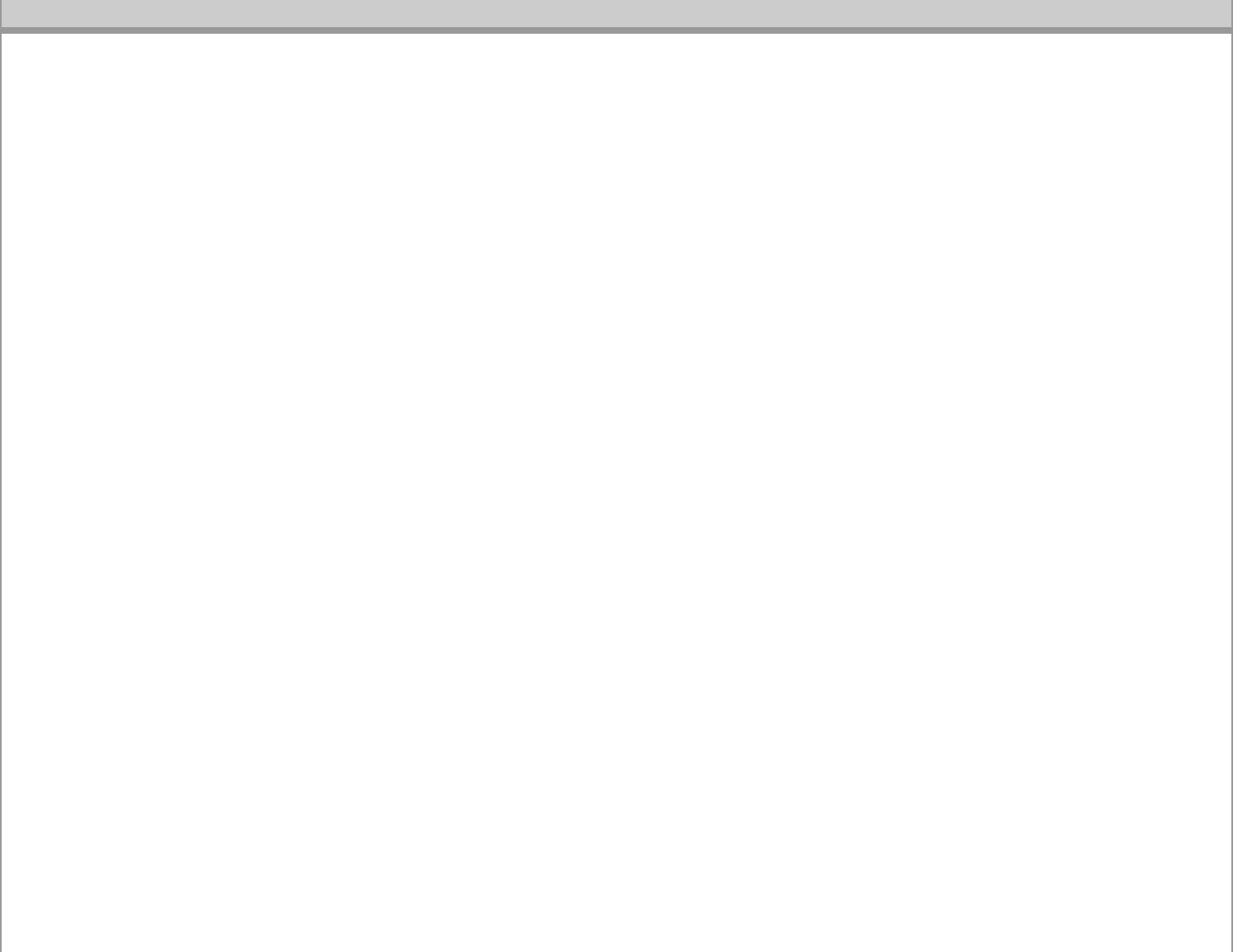
The duplication of a living cell *in silico* represents an extremely formidable undertaking. Not only must the virtual cell possess all of the proteins and metabolites for the type of cell to be modeled (eg, from brain, liver, nerve, muscle, or adipose), but these must be present in the appropriate concentration and subcellular location. The model must also account for the functional dynamics of its components, binding affinities, catalytic efficiency, covalent modifications, etc. To render a virtual cell capable of dividing or differentiating will entail a further quantum leap in complexity and sophistication.

### Molecular Interaction Maps Employ Symbolic Logic

The models constructed by systems biologists can take a variety of forms depending upon the uses for which they are intended and the data available to guide their construction. If one wishes to model the flux of metabolites through an anabolic or catabolic pathway, it is not enough to know the identities and the reactants involved in each enzyme-catalyzed reaction. To obtain mathematically precise values, it is necessary to know the concentrations of the metabolites in question, the quantity of each of each enzyme present, and their catalytic parameters.

For most users, it is sufficient that a model describe and predict the qualitative nature of the interactions between components. Does an allosteric ligand activate or inhibit the enzyme? Does dissociation of a protein complex lead to the degradation of one or more of its components? For this purpose, a set of symbols depicting the symbolic logic of these interactions was needed. Early representations frequently used the symbols previously developed for constructing flow charts (computer programming) or electronic circuits (**Figure 10–8**, top). Ultimately, however, systems biologists designed dedicated symbols (**Figure 10–8**, bottom) to depict these molecular circuit diagrams, more commonly referred to as **Molecular Interaction Maps (MIM)**, an example of which is shown in **Figure 10–9**. Unfortunately, as is the case with enzyme nomenclature (Chapter 7) a consistent, universal set of symbols has yet to emerge.

**Figure 10–8**



	Process	An operation or action step
	Terminator	A start or stop point in a process
	Decision	A question or branch in a process
	Sort	Sorting into some predetermined order
	Merge	Merge multiple processes into one
	Data	Indicates data inputs to or from a process
	Connector	A jump from one point to another

#### Reaction symbols

- (a) Non-covalent binding (reversible)
- (b) Covalent modification.
- (b') Covalent bond (see Figure 13 & text).
- (c) Stoichiometric conversion
- (d) Products appear without loss of reactants.
- (e) Transcription
- (f) Cleavage of a covalent bond.
- (g) Degradation
- (h) Reaction *in trans*.

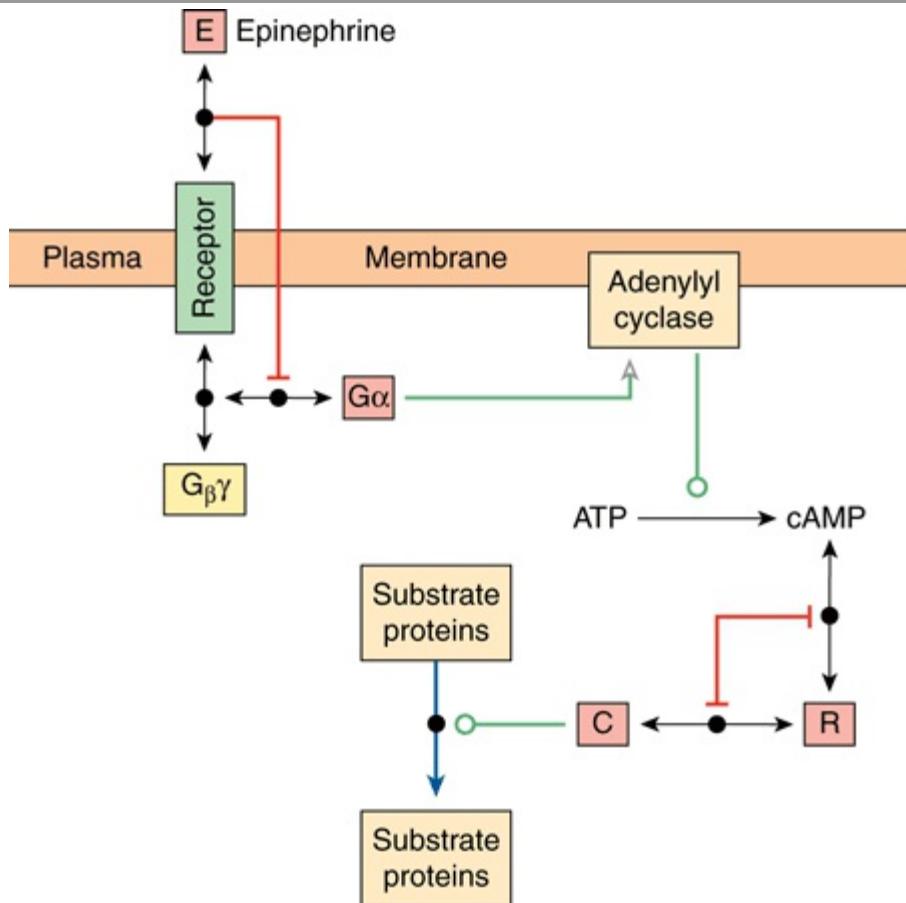
#### Contingency symbols

- (i) Stimulation
- (j) Requirement
- (k) Inhibition
- (l) Enzymatic catalysis

Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Figure 10–9**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Representation of a molecular interaction network (MIN) depicting a signal transduction cascade leading to the phosphorylation of substrate proteins by the catalytic subunit, C, of the cyclic AMP-dependent protein kinase in response to epinephrine.** Proteins are depicted as rectangles or squares. Double headed arrows indicate formation of noncovalent complex represented by dot at the midpoint of the arrow. Red lines with T-shaped heads indicate inhibitory interaction. Green arrow with hollow head indicates a stimulatory interaction. Green line with open circle at end indicates catalysis. Blue arrow with P indicates covalent modification by phosphorylation. (Symbols adapted from Kohn KW et al: Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. Mol Biol Cell 2006; 17:1.)

## CONCLUSION

The rapidly evolving fields of bioinformatics and computational biology hold unparalleled promise for the future of both medicine and basic biology. Some promises are at present perceived clearly, others dimly, while yet others remain unimagined. A major objective of computational biologists is to develop computational tools that will enhance the efficiency, effectiveness, and speed of drug development. Epidemiologists employ computers to extract patterns within a population indicative of specific causes of and contributors to both disease and wellness. There seems little doubt that their impact on medical practice in the 21st century will equal or surpass that of the discovery of bacterial pathogenesis in the 19th century.

## SUMMARY

- Genomics has yielded a massive quantity of information of great potential value to scientists and physicians.
- Bioinformatics involves the design of computer algorithms and construction of databases that enable biomedical scientists to access and analyze the growing avalanche of biomedical data.
- The objective of epidemiology is to extract medical insights from the behavior of heterogeneous human populations by the application of sophisticated statistical tools.
- Major challenges in the construction of user-friendly databases include devising means for storing and organizing complex data that accommodate a wide range of potential search criteria.
- The goal of the Encode Project is to identify all the functional elements within the human genome.
- The HapMap, Entrez Gene, and dbGAP databases contain data concerning the relation of genetic mutations to pathological conditions.
- Computational biology uses computer algorithms to identify unknown proteins and conduct virtual experiments.
- BLAST is used to identify unknown proteins and genes by searching for sequence homologs of known function.
- Computational biologists are developing programs that will predict the three-dimensional structure of proteins directly from their primary sequence.
- Computer-aided drug design speeds drug discovery by trying to dock potential inhibitors to selected protein targets *in silico*.
- A major goal of systems biologists is to create faithful models of individual pathways and networks in order to elucidate functional principles and perform virtual experiments.
- The ultimate goal of systems biologists is to create virtual cells that can be used to more safely and efficiently diagnose and treat diseases, particularly those of a multifactorial nature.
- Systems biologists commonly construct schematic representations known as molecular interaction maps in which symbolic logic is employed to illustrate the relationships between the components making up a pathway or some other functional unit

## REFERENCES

- Altschul SF, Gish W, Miller W, et al: Basic local alignment search tool. *J Mol Biol* 1990;215:403.
- Collins FS, Barker AD: Mapping the human cancer genome. *Sci Am* 2007;296:50.
- Collins FS, Green ED, Guttmacher AE, et al: A vision for the future of genomics research. A blueprint for the genomic era. *Nature* 2003;422:835.
- Couzin J: The HapMap gold rush: researchers mine a rich deposit. *Science* 2006;312:1131.
- Cravatt BF, Wright AT, Kozarich JW: Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu Rev Biochem* 2008;77:383.
- Debes JD, Urrutia R: Bioinformatics tools to understand human diseases. *Surgery* 2004;135:579.
- Dunning Hotopp JC, Griffantini R, Kumar N, et al: Comparative genomics of *Neisseria meningitidis*: core genome, islands of horizontal transfer and pathogen-specific genes. *Microbiology* 2006;152:3691.
- Ekins S, Mestres J, Testa B: *In silico* pharmacology for drug discovery: applications to targets and beyond. *Br J Pharmacol* 2007;152:21.
- Ekins S, Mestres J, Testa B: *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol* 2007;152:9.
- Kaiser J: Affordable 'exomes' fill gaps in a catalog of rare diseases. *Science* 2010;330:903.
- Kim JH: Bioinformatics and genomic medicine. *Genet Med* 2002;4:62S.
- Kohn KW, Aladjem MI, Weinstein JN, et al: Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Mol Biol Cell* 2006; 17:1.
- Koonin EV, Galperin MY: *Sequence—Evolution—Function. Computational Approaches to Comparative Genomics*. Kluwer Academic Publishers, 2003.
- Laurie ATR, Jackson RM: Methods for prediction of protein–ligand binding sites for structure-based drug design and virtual ligand screening. *Curr Prot Peptide Sci* 2006; 7:395–406.
- McInnes C: Virtual screening strategies in drug discovery. *Curr Opin Cell Biol* 2007;11:494.
- Nebert DW, Zhang G, Vesell ES: From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metab Rev* 2008; 40:187.
- Sansom C: Genes and disease. *The Scientist* 2008;30:34.
- Slepchenko BM, Schaff JC, Macara I, et al: Quantitative cell biology with the Virtual Cell. *Trends Cell Biol* 2003; 13:570.
- Sudmant PH, Kitzman JO, Antonacci F, et al: Diversity of human gene copy number variation and multicopy genes. *Science* 2010; 330:641.
- Villoutreix BO, Renault N, Lagorce D, et al: Free resources to assist structure-based virtual ligand screening experiments. *Curr Protein Pept Sci* 2007;8:381.

## OBJECTIVES

After studying this chapter, you should be able to:

- State the first and second laws of thermodynamics and understand how they apply to biologic systems.
- Explain what is meant by the terms free energy, entropy, enthalpy, exergonic, and endergonic.
- Appreciate how reactions that are endergonic may be driven by coupling to those that are exergonic in biologic systems.
- Understand the role of high-energy phosphates, ATP, and other nucleotide triphosphates in the transfer of free energy from exergonic to endergonic processes, enabling them to act as the "energy currency" of cells.

## BIOMEDICAL IMPORTANCE

Bioenergetics, or biochemical thermodynamics, is the study of the energy changes accompanying biochemical reactions. Biologic systems are essentially **isothermal** and use chemical energy to power living processes. How an animal obtains suitable fuel from its food to provide this energy is basic to the understanding of normal nutrition and metabolism. Death from **starvation** occurs when available energy reserves are depleted, and certain forms of malnutrition are associated with energy imbalance (**marasmus**). Thyroid hormones control the rate of energy release (metabolic rate), and disease results when they malfunction. Excess storage of surplus energy causes **obesity**, an increasingly common disease of Western society, which predisposes to many diseases, including cardiovascular disease and diabetes mellitus type 2, and lowers life expectancy.

## FREE ENERGY IS THE USEFUL ENERGY IN A SYSTEM

Gibbs change in free energy ( $\Delta G$ ) is that portion of the total energy change in a system that is available for doing work—ie, the useful energy, also known as the chemical potential.

### BioLogic Systems Conform to the General Laws of Thermodynamics

The first law of thermodynamics states that **the total energy of a system, including its surroundings, remains constant**. It implies that within the total system, energy is neither lost nor gained during any change. However, energy may be transferred from one part of the system to another or may be transformed into another form of energy. In living systems, chemical energy may be transformed into heat or into electrical, radiant, or mechanical energy.

The second law of thermodynamics states that **the total entropy of a system must increase if a process is to occur spontaneously**. **Entropy** is the extent of disorder or randomness of the system and becomes maximum as equilibrium is approached. Under conditions of constant temperature and pressure, the relationship between the free-energy change ( $\Delta G$ ) of a reacting system and the change in entropy ( $\Delta S$ ) is expressed by the following equation, which combines the two laws of thermodynamics:

$$\Delta G = \Delta H - T\Delta S$$

where  $\Delta H$  is the change in **enthalpy** (heat) and  $T$  is the absolute temperature.

In biochemical reactions, since  $\Delta H$  is approximately equal to  $\Delta E$ , the total change in internal energy of the reaction, the above relationship may be expressed in the following way:

$$\Delta G = \Delta E - T\Delta S$$

If  $\Delta G$  is negative, the reaction proceeds spontaneously with loss of free energy; ie, it is **exergonic**. If, in addition,  $\Delta G$  is of great magnitude, the reaction goes virtually to completion and is essentially irreversible. On the other hand, if  $\Delta G$  is positive, the reaction proceeds only if free energy can be gained; ie, it is **endergonic**. If, in addition, the magnitude of  $\Delta G$  is great, the system is stable, with little or no tendency for a reaction to occur. If  $\Delta G$  is zero, the system is at equilibrium and no net change takes place.

When the reactants are present in concentrations of 1.0 mol/L,  $\Delta G^0$  is the standard free-energy change. For biochemical reactions, a standard state is defined as having a pH of 7.0. The standard free-energy change at this standard state is denoted by  $\Delta G^0'$ .

The standard free-energy change can be calculated from the equilibrium constant  $K_{eq}$ .

$$\Delta G^0' = -RT \ln K'_{eq}$$

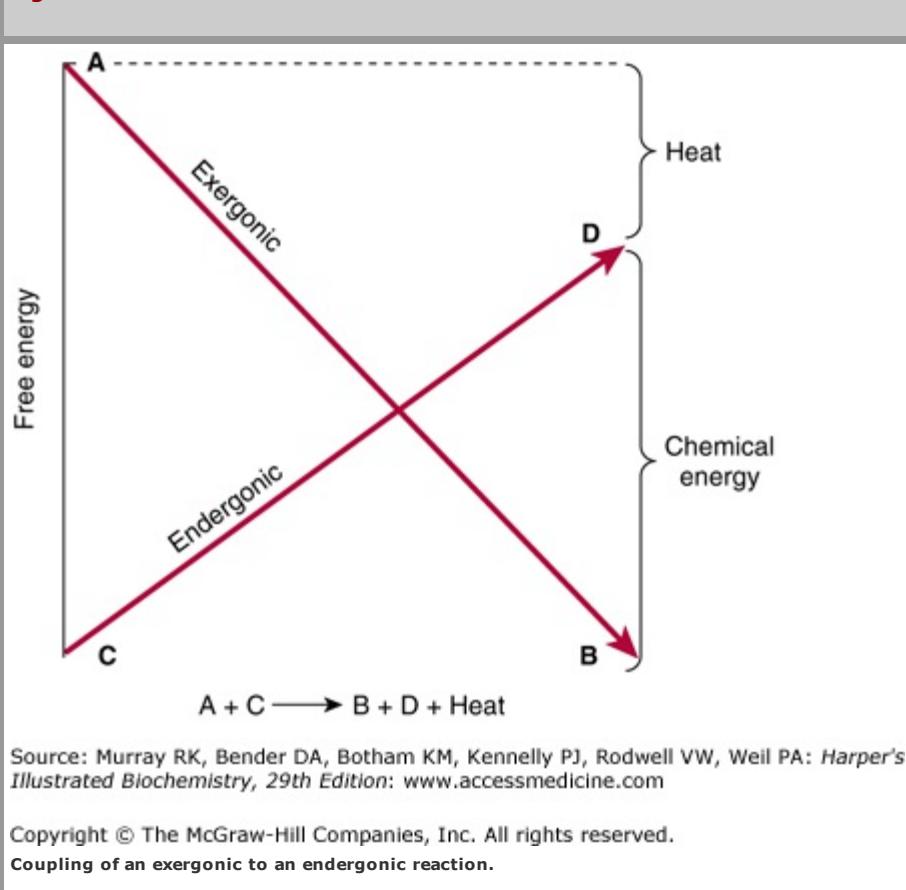
where  $R$  is the gas constant and  $T$  is the absolute temperature (Chapter 8). It is important to note that the actual  $\Delta G$  may be larger or smaller than  $\Delta G^0'$  depending on the concentrations of the various reactants, including the solvent, various ions, and proteins.

In a biochemical system, an enzyme only speeds up the attainment of equilibrium; it never alters the final concentrations of the reactants at equilibrium.

## ENDERGONIC PROCESSES PROCEED BY COUPLING TO EXERGONIC PROCESSES

The vital processes—eg, synthetic reactions, muscular contraction, nerve impulse conduction, and active transport—obtain energy by chemical linkage, or **coupling**, to oxidative reactions. In its simplest form, this type of coupling may be represented as shown in **Figure 11-1**. The conversion of metabolite A to metabolite B occurs with release of free energy and is coupled to another reaction in which free energy is required to convert metabolite C to metabolite D. The terms **exergonic** and **endergonic**, rather than the normal chemical terms "exothermic" and "endothermic," are used to indicate that a process is accompanied by loss or gain, respectively, of free energy in any form, not necessarily as heat. In practice, an endergonic process cannot exist independently, but must be a component of a coupled exergonic-endergonic system where the overall net change is exergonic. The exergonic reactions are termed **catabolism** (generally, the breakdown or oxidation of fuel molecules), whereas the synthetic reactions that build up substances are termed **anabolism**. The combined catabolic and anabolic processes constitute **metabolism**.

**Figure 11-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

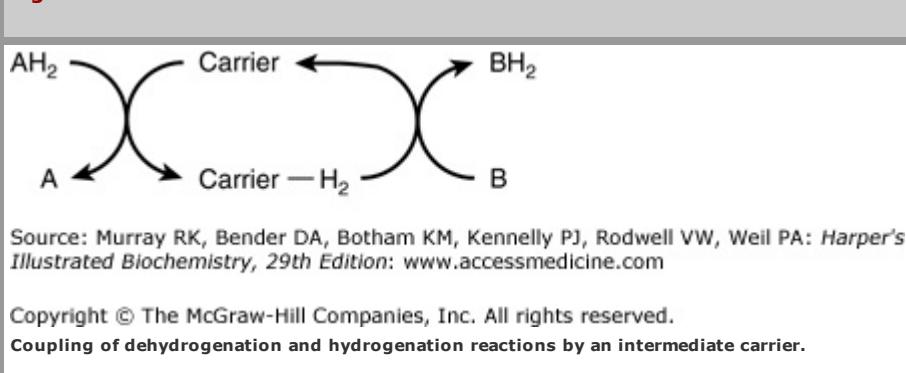
Coupling of an exergonic to an endergonic reaction.

If the reaction shown in **Figure 11-1** is to go from left to right, then the overall process must be accompanied by loss of free energy as heat. One possible mechanism of coupling could be envisaged if a common obligatory intermediate (I) took part in both reactions, ie,



Some exergonic and endergonic reactions in biologic systems are coupled in this way. This type of system has a built-in mechanism for biologic control of the rate of oxidative processes since the common obligatory intermediate allows the rate of utilization of the product of the synthetic path (D) to determine by mass action the rate at which A is oxidized. Indeed, these relationships supply a basis for the concept of **respiratory control**, the process that prevents an organism from burning out of control. An extension of the coupling concept is provided by dehydrogenation reactions, which are coupled to hydrogenations by an intermediate carrier (**Figure 11-2**).

**Figure 11-2**



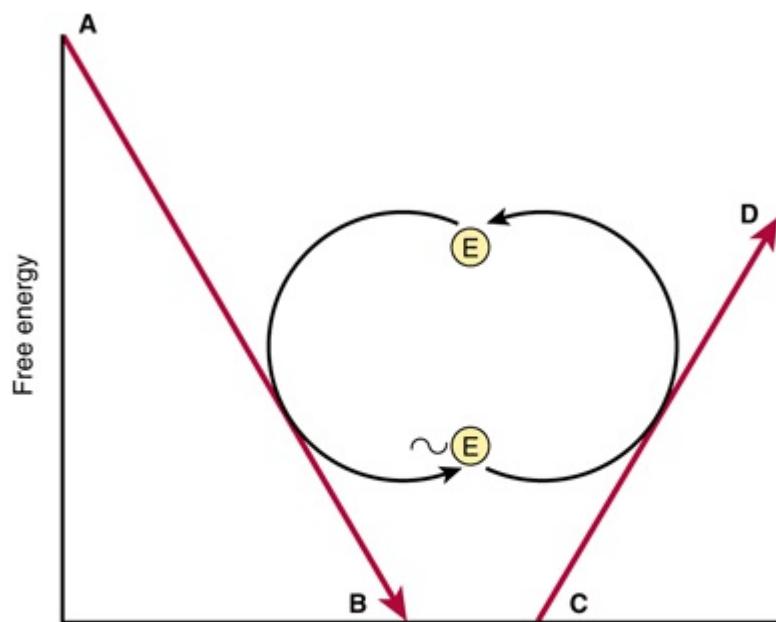
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coupling of dehydrogenation and hydrogenation reactions by an intermediate carrier.

An alternative method of coupling an exergonic to an endergonic process is to synthesize a compound of high-energy potential in the exergonic reaction and to incorporate this new compound into the endergonic reaction, thus effecting a transference of free energy from the exergonic pathway (**Figure 11-3**). The biologic advantage of this mechanism is that the compound of high potential energy, ~E, unlike I in the previous system, need not be structurally related to A, B, C, or D, allowing E to serve as a transducer of energy from a wide range of exergonic reactions or processes, such as biosyntheses, muscular contraction, nervous excitation, and active transport. In the living cell, the principal high-energy intermediate or carrier compound (designated ~E in **Figure 11-3**) is **adenosine triphosphate (ATP)** (**Figure 11-4**).

**Figure 11-3**

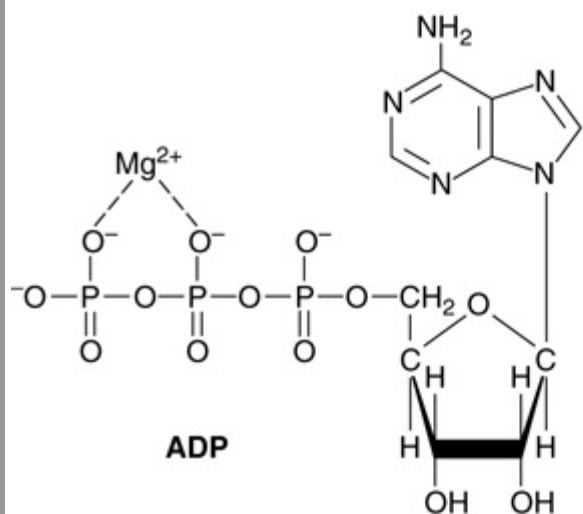
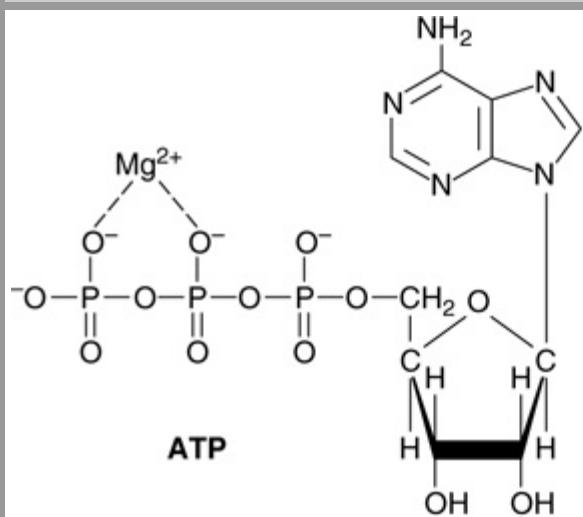


Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transfer of free energy from an exergonic to an endergonic reaction via a high-energy intermediate compound ( $\sim E$ ).

**Figure 11–4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adenosine triphosphate (ATP) and adenosine diphosphate shown as the magnesium complexes.



## HIGH-ENERGY PHOSPHATES PLAY A CENTRAL ROLE IN ENERGY CAPTURE AND TRANSFER

In order to maintain living processes, all organisms must obtain supplies of free energy from their environment. **Autotrophic** organisms utilize simple exergonic processes; eg, the energy of sunlight (green plants), the reaction  $\text{Fe}^{2+} \rightarrow \text{Fe}^{3+}$  (some bacteria). On the other hand, **heterotrophic** organisms obtain free energy by coupling their metabolism to the breakdown of complex organic molecules in their environment. In all these organisms, ATP plays a central role in the transference of free energy from the exergonic to the endergonic processes (**Figure 11–3**). ATP is a nucleoside triphosphate containing adenine, ribose, and three phosphate groups. In its reactions in the cell, it functions as the  $\text{Mg}^{2+}$  complex (**Figure 11–4**).

The importance of phosphates in intermediary metabolism became evident with the discovery of the role of ATP, adenosine diphosphate (ADP) (**Figure 11–4**), and inorganic phosphate ( $\text{P}_i$ ) in glycolysis (Chapter 18).

### The Intermediate Value for the Free Energy of Hydrolysis of ATP Has Important Bioenergetic Significance

The standard free energy of hydrolysis of a number of biochemically important phosphates is shown in **Table 11–1**. An estimate of the comparative tendency of each of the phosphate groups to transfer to a suitable acceptor may be obtained from the  $\Delta G^0$  of hydrolysis at 37°C. The value for the hydrolysis of the terminal phosphate of ATP divides the list into two groups. **Low-energy phosphates**, exemplified by the ester phosphates found in the intermediates of glycolysis, have  $G^0$  values smaller than that of ATP, while in **high-energy phosphates** the value is higher than that of ATP. The components of this latter group, including ATP, are usually anhydrides (eg, the 1-phosphate of 1,3-bisphosphoglycerate), enolphosphates (eg, phosphoenolpyruvate), and phosphoguanidines (eg, creatine phosphate, arginine phosphate).

**Table 11–1 Standard Free Energy of Hydrolysis of Some Organophosphates of Biochemical Importance**

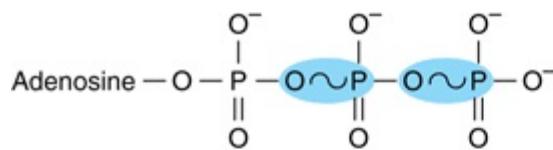
<b>Compound</b>	$\Delta G^0$	
	<b>kJ/mol</b>	<b>kcal/mol</b>
Phosphoenolpyruvate	-61.9	-14.8
Carbamoyl phosphate	-51.4	-12.3
1,3-Bisphosphoglycerate (to 3-phosphoglycerate)	-49.3	-11.8
Creatine phosphate	-43.1	-10.3
ATP $\rightarrow$ AMP + PP <i>i</i>	-32.2	-7.7
ATP $\rightarrow$ ADP + $\text{P}_i$	-30.5	-7.3
Glucose 1-phosphate	-20.9	-5.0
PP <i>i</i>	-19.2	-4.6
Fructose 6-phosphate	-15.9	-3.8
Glucose 6-phosphate	-13.8	-3.3
Glycerol 3-phosphate	-9.2	-2.2

**Abbreviations:** PP*i*, pyrophosphate;  $\text{P}_i$ , inorganic orthophosphate.

**Note:** All values taken from Jencks (1976), except that for PP*i* which is from Frey and Arabshahi (1995). Values differ between investigators, depending on the precise conditions under which the measurements were made.

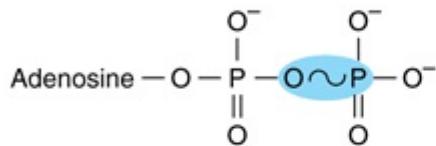
The symbol  $\sim \text{P}$  indicates that the group attached to the bond, on transfer to an appropriate acceptor, results in transfer of the larger quantity of free energy. For this reason, the term **group transfer potential**, rather than "high-energy bond," is preferred by some. Thus, ATP contains two high-energy phosphate groups and ADP contains one, whereas the phosphate in AMP (adenosine monophosphate) is of the low-energy type since it is a normal ester link (**Figure 11–5**).

**Figure 11–5**



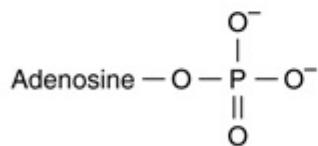
or Adenosine — (P) ~ (P) ~ (P)

### Adenosine triphosphate (ATP)



or Adenosine — (P) ~ (P)

### Adenosine diphosphate (ADP)



or Adenosine — (P)

### Adenosine monophosphate (AMP)

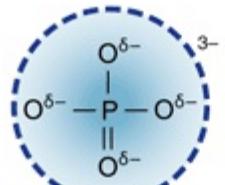
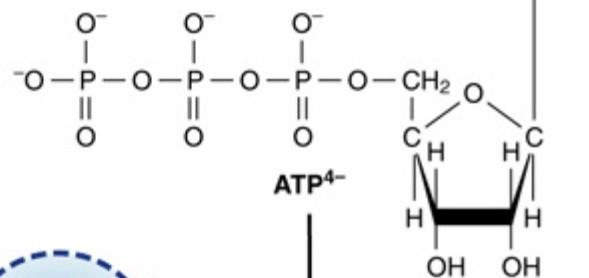
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Structure of ATP, ADP, and AMP showing the position and the number of high-energy phosphates (~P).**

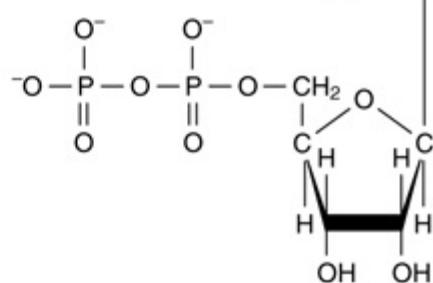
The intermediate position of ATP allows it to play an important role in energy transfer. The high free-energy change on hydrolysis of ATP is due to relief of charge repulsion of adjacent negatively charged oxygen atoms and to stabilization of the reaction products, especially phosphate, as resonance hybrids (**Figure 11–6**). Other "high-energy compounds" are thiol esters involving coenzyme A (eg, acetyl-CoA), acyl carrier protein, amino acid esters involved in protein synthesis, S-adenosylmethionine (active methionine), UDPGlc (uridine diphosphate glucose), and PRPP (5-phosphoribosyl-1-pyrophosphate).

**Figure 11–6**



The phosphate released is stabilised by the formation of a resonance hybrid in which the 3 negative charges are shared between the four O atoms

Hydrolysis of  $\text{ATP}^{4-}$  to  $\text{ADP}^{3-}$  relieves charge repulsion



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

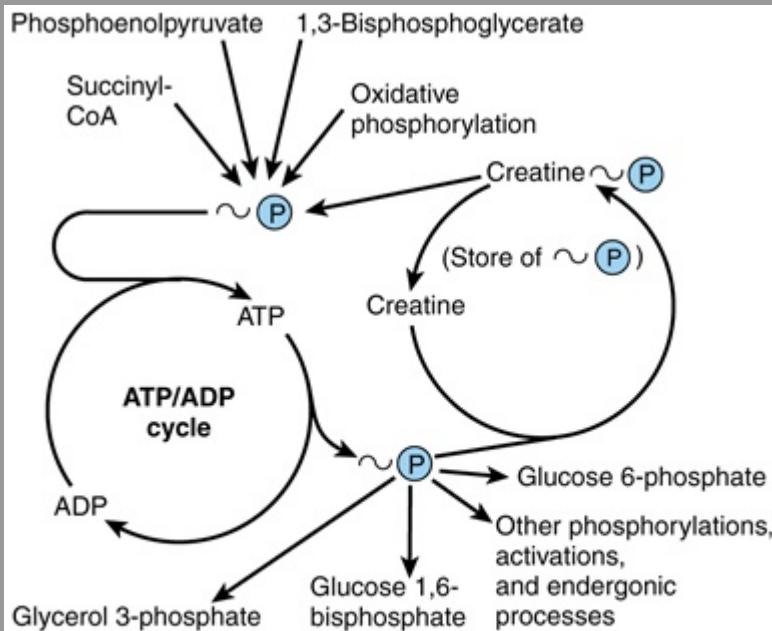
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The free-energy change on hydrolysis of ATP to ADP.

## HIGH-ENERGY PHOSPHATES ACT AS THE "ENERGY CURRENCY" OF THE CELL

ATP is able to act as a donor of high-energy phosphate to form those compounds below it in **Table 11–1**. Likewise, with the necessary enzymes, ADP can accept high-energy phosphate to form ATP from those compounds above ATP in the table. In effect, an **ATP/ADP cycle** connects those processes that generate  $\sim\text{P}$  to those processes that utilize  $\sim\text{P}$  (Figure 11–7), continuously consuming and regenerating ATP. This occurs at a very rapid rate since the total ATP/ADP pool is extremely small and sufficient to maintain an active tissue for only a few seconds.

**Figure 11–7**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

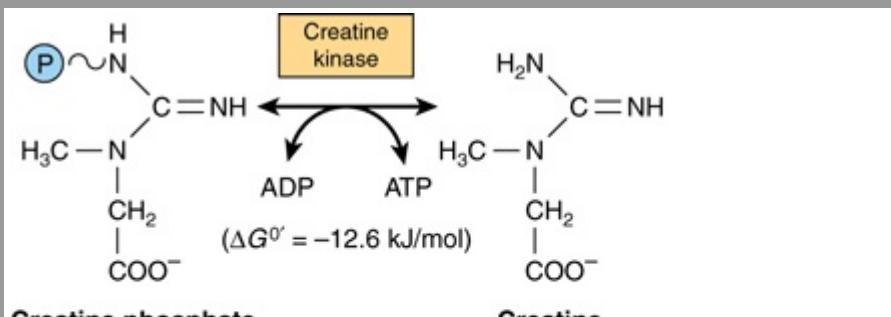
**Role of ATP/ADP cycle in transfer of high-energy phosphate.**

There are three major sources of  $\sim\text{P}$  taking part in **energy conservation** or **energy capture**:

1. **Oxidative phosphorylation.** The greatest quantitative source of  $\sim\text{P}$  in aerobic organisms. Free energy comes from respiratory chain oxidation using molecular O<sub>2</sub> within mitochondria (Chapter 12).
2. **Glycolysis.** A net formation of two  $\sim\text{P}$  results from the formation of lactate from one molecule of glucose, generated in two reactions catalyzed by phosphoglycerate kinase and pyruvate kinase, respectively (Figure 18–2).
3. **The citric acid cycle.** One  $\sim\text{P}$  is generated directly in the cycle at the succinate thiokinase step (Figure 17–3).

**Phosphagens** act as storage forms of high-energy phosphate and include creatine phosphate, which occurs in vertebrate skeletal muscle, heart, spermatozoa, and brain, and arginine phosphate, which occurs in invertebrate muscle. When ATP is rapidly being utilized as a source of energy for muscular contraction, phosphagens permit its concentrations to be maintained, but when the ATP/ADP ratio is high, their concentration can increase to act as a store of high-energy phosphate (Figure 11–8).

**Figure 11–8**



**Creatine phosphate**

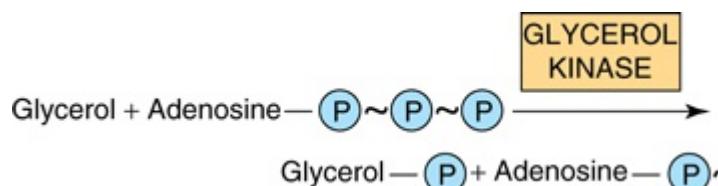
**Creatine**

Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

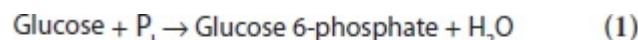
**Transfer of high-energy phosphate between ATP and creatine.**

When ATP acts as a phosphate donor to form those compounds of lower free energy of hydrolysis (**Table 11–1**), the phosphate group is invariably converted to one of low energy, eg



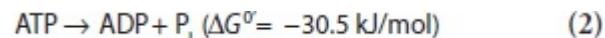
### ATP Allows the Coupling of Thermodynamically Unfavorable Reactions to Favorable Ones

The phosphorylation of glucose to glucose 6-phosphate, the first reaction of glycolysis (Figure 18–2), is highly endergonic and cannot proceed under physiologic conditions:



$$(\Delta G^\circ = +13.8 \text{ kJ/mol})$$

To take place, the reaction must be coupled with another—more exergonic—reaction such as the hydrolysis of the terminal phosphate of ATP.



When (1) and (2) are coupled in a reaction catalyzed by hexokinase, phosphorylation of glucose readily proceeds in a highly exergonic reaction that under physiologic conditions is irreversible. Many "activation" reactions follow this pattern.

### Adenylyl Kinase (Myokinase) Interconverts Adenine Nucleotides

This enzyme is present in most cells. It catalyzes the following reaction:

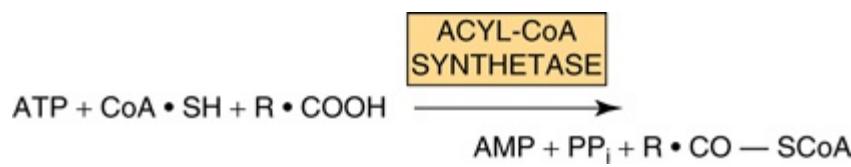


This allows:

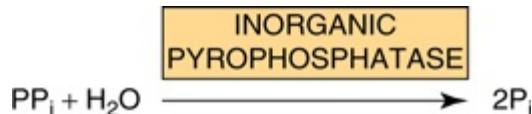
1. High-energy phosphate in ADP to be used in the synthesis of ATP.
2. AMP, formed as a consequence of several activating reactions involving ATP, to be recovered by rephosphorylation to ADP.
3. AMP to increase in concentration when ATP becomes depleted and act as a metabolic (allosteric) signal to increase the rate of catabolic reactions, which in turn lead to the generation of more ATP (Chapter 20).

### When ATP Forms AMP, Inorganic Pyrophosphate ( $\text{PP}_i$ ) Is Produced

ATP can also be hydrolyzed directly to AMP, with the release of  $\text{PP}_i$  (Table 11–1). This occurs, for example, in the activation of long-chain fatty acids (Chapter 22).

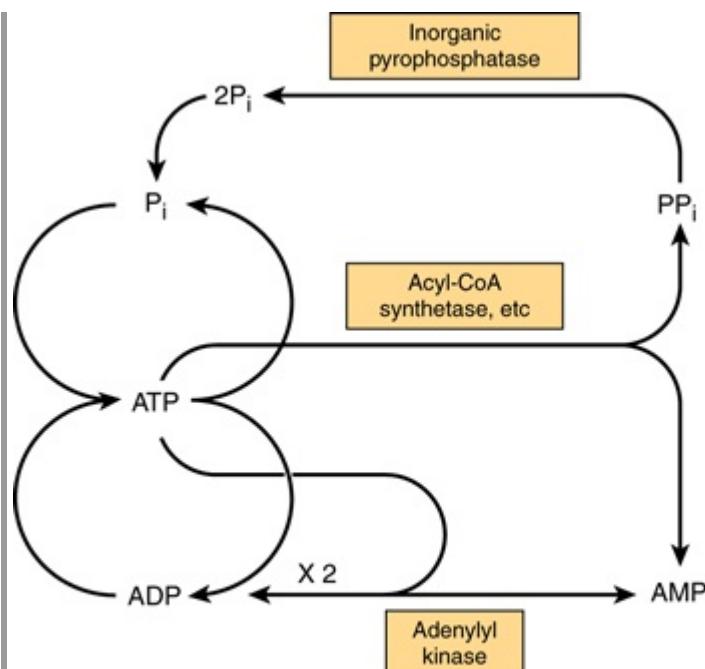


This reaction is accompanied by loss of free energy as heat, which ensures that the activation reaction will go to the right and is further aided by the hydrolytic splitting of  $\text{PP}_i$ , catalyzed by **inorganic pyrophosphatase**, a reaction that itself has a large  $\Delta G^\circ$  of  $-19.2 \text{ kJ/mol}$ . Note that activations via the pyrophosphate pathway result in the loss of two  $\sim \text{P}$  rather than one, as occurs when ADP and  $\text{P}_i$  are formed.



A combination of the above reactions makes it possible for phosphate to be recycled and the adenine nucleotides to interchange (Figure 11–9).

**Figure 11–9**



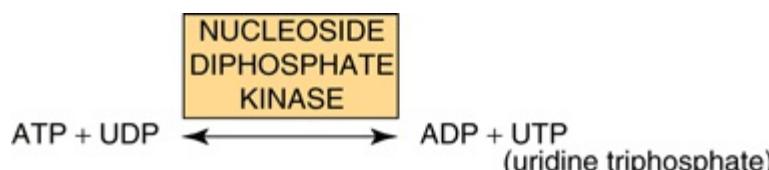
Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

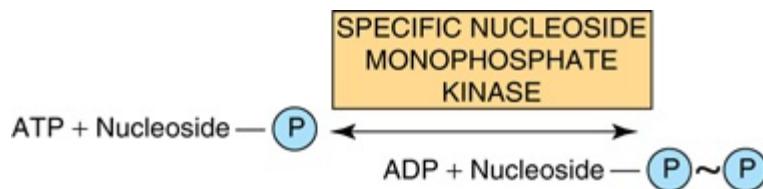
Phosphate cycles and interchange of adenine nucleotides.

### Other Nucleoside Triphosphates Participate in the Transfer of High-Energy Phosphate

By means of the enzyme **nucleoside diphosphate kinase**, UTP, GTP, and CTP can be synthesized from their diphosphates, eg, UDP reacts with ATP to form UTP.



All of these triphosphates take part in phosphorylations in the cell. Similarly, specific nucleoside monophosphate kinases catalyze the formation of nucleoside diphosphates from the corresponding monophosphates.



Thus, adenylyl kinase is a specialized monophosphate kinase.

## SUMMARY

- Biologic systems use chemical energy to power living processes.
- Exergonic reactions take place spontaneously with loss of free energy ( $\Delta G$  is negative). Endergonic reactions require the gain of free energy ( $\Delta G$  is positive) and occur only when coupled to exergonic reactions.
- ATP acts as the "energy currency" of the cell, transferring free energy derived from substances of higher energy potential to those of lower energy potential.

## REFERENCES

- de Meis L: The concept of energy-rich phosphate compounds: water, transport ATPases, and entropy energy. *Arch Biochem Biophys* 1993;306:287.
- Frey PA, Arabshahi A: Standard free-energy change for the hydrolysis of the alpha, beta-phosphoanhydride bridge in ATP. *Biochemistry* 1995;34:11307.
- Harold FM: *The Vital Force: A Study of Bioenergetics*. Freeman, 1986.
- Harris DA: *Bioenergetics at a Glance: An Illustrated Introduction*. Blackwell Publishing, 1995.
- Haynie D: *Biological Thermodynamics*. Cambridge University Press, 2008.
- Jencks WP: Free energies of hydrolysis and decarboxylation. In: *Handbook of Biochemistry and Molecular Biology*, vol 1. *Physical and Chemical Data*. Fasman GD (editor). CRC Press, 1976:296–304.
- Klotz IM: *Introduction to Biomolecular Energetics*. Academic Press, 1986.
- Nicholls D, Ferguson F: *Bioenergetics*. Elsevier, 2003.

## OBJECTIVES

After studying this chapter, you should be able to:

- Understand the meaning of redox potential and explain how it can be used to predict the direction of flow of electrons in biologic systems.
- Identify the four classes of enzymes (oxidoreductases) involved in oxidation and reduction reactions.
- Describe the action of oxidases and provide examples of where they play an important role in metabolism.
- Indicate the two main functions of dehydrogenases and explain the importance of NAD- and riboflavin-linked dehydrogenases in metabolic pathways such as glycolysis, the citric acid cycle, and the respiratory chain.
- Identify the two types of enzymes classified as hydroperoxidases; indicate the reactions they catalyze and explain why they are important.
- Give the two steps of reactions catalyzed by oxygenases and identify the two subgroups of this class of enzymes.
- Appreciate the role of cytochrome P450 in drug detoxification and steroid synthesis.
- Describe the reaction catalyzed by superoxide dismutase and explain how it protects tissues from oxygen toxicity.

## BIOMEDICAL IMPORTANCE

Chemically, **oxidation** is defined as the removal of electrons and **reduction** as the gain of electrons. Thus, oxidation is always accompanied by reduction of an electron acceptor. This principle of oxidation-reduction applies equally to biochemical systems and is an important concept underlying understanding of the nature of biologic oxidation. Note that many biologic oxidations can take place without the participation of molecular oxygen, eg, dehydrogenations. The life of higher animals is absolutely dependent upon a supply of oxygen for **respiration**, the process by which cells derive energy in the form of ATP from the controlled reaction of hydrogen with oxygen to form water. In addition, molecular oxygen is incorporated into a variety of substrates by enzymes designated as **oxygenases**; many drugs, pollutants, and chemical carcinogens (xenobiotics) are metabolized by enzymes of this class, known as the **cytochrome P450 system**. Administration of oxygen can be lifesaving in the treatment of patients with respiratory or circulatory failure.

## FREE ENERGY CHANGES CAN BE EXPRESSED IN TERMS OF REDOX POTENTIAL

In reactions involving oxidation and reduction, the free energy change is proportionate to the tendency of reactants to donate or accept electrons. Thus, in addition to expressing free energy change in terms of  $\Delta G^{\circ}$  (Chapter 11), it is possible, in an analogous manner, to express it numerically as an **oxidation-reduction** or **redox potential** ( $E^{\circ}$ ). The redox potential of a system ( $E_0$ ) is usually compared with the potential of the hydrogen electrode (0.0 V at pH 0.0). However, for biologic systems, the redox potential ( $E^{\circ}_0$ ) is normally expressed at pH 7.0, at which pH the electrode potential of the hydrogen electrode is -0.42 V. The redox potentials of some redox systems of special interest in mammalian biochemistry are shown in **Table 12-1**. The relative positions of redox systems in the table allow prediction of the direction of flow of electrons from one redox couple to another.

**Table 12-1 Some Redox Potentials of Special Interest in Mammalian Oxidation Systems**

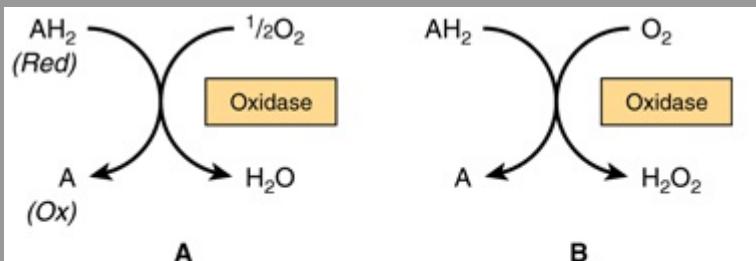
System	$E^{\circ}_0$ Volts
H <sup>+</sup> /H <sub>2</sub>	- 0.42
NAD <sup>+</sup> /NADH	- 0.32
Lipoate; ox/red	- 0.29
Acetoacetate/3-hydroxybutyrate	- 0.27
Pyruvate/lactate	- 0.19
Oxaloacetate/malate	- 0.17
Fumarate/succinate	+0.03
Cytochrome <i>b</i> ; Fe <sup>3+</sup> /Fe <sup>2+</sup>	+0.08
Ubiquinone; ox/red	+0.10
Cytochrome <i>c</i> <sub>1</sub> ; Fe <sup>3+</sup> /Fe <sup>2+</sup>	+0.22
Cytochrome <i>a</i> ; Fe <sup>3+</sup> /Fe <sup>2+</sup>	+0.29
Oxygen/water	+0.82

Enzymes involved in oxidation and reduction are called **oxidoreductases** and are classified into four groups: **oxidases**, **dehydrogenases**, **hydroperoxidases**, and **oxygenases**.

## OXIDASES USE OXYGEN AS A HYDROGEN ACCEPTOR

Oxidases catalyze the removal of hydrogen from a substrate using oxygen as a hydrogen acceptor.\* They form water or hydrogen peroxide as a reaction product (Figure 12-1).

**Figure 12-1**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oxidation of a metabolite catalyzed by an oxidase (A) forming  $H_2O$  and (B) forming  $H_2O_2$ .

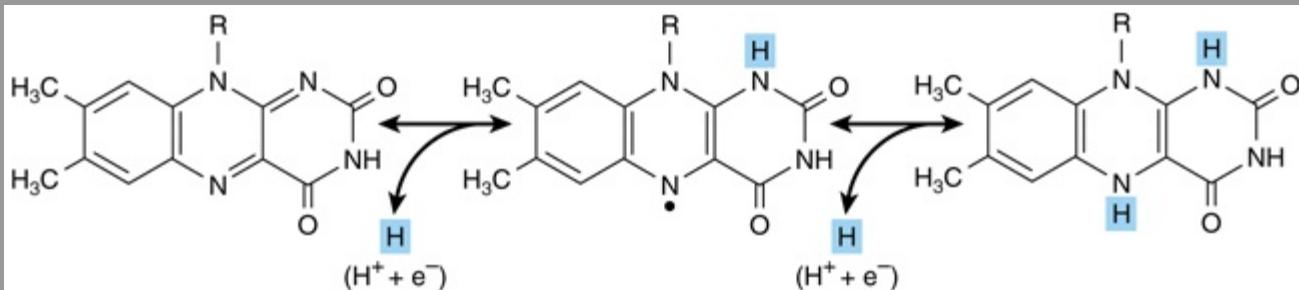
## Some Oxidases Contain Copper

**Cytochrome oxidase** is a hemoprotein widely distributed in many tissues, having the typical heme prosthetic group present in myoglobin, hemoglobin, and other cytochromes (Chapter 6). It is the terminal component of the chain of respiratory carriers found in mitochondria (Chapter 13) and transfers electrons resulting from the oxidation of substrate molecules by dehydrogenases to their final acceptor, oxygen. The action of the enzyme is blocked by carbon monoxide, cyanide, and hydrogen sulfide, and this causes poisoning by preventing cellular respiration. It has also been termed "cytochrome  $a_3$ ." However, it is now known that the heme  $a_3$  is combined with another heme, heme  $a$ , in a single protein to form the cytochrome oxidase enzyme complex, and so it is more correctly termed **cytochrome  $a_{a3}$** . It contains two molecules of heme, each having one Fe atom that oscillates between  $Fe^{3+}$  and  $Fe^{2+}$  during oxidation and reduction. Furthermore, two atoms of Cu are present, each associated with a heme unit.

## Other Oxidases Are Flavoproteins

Flavoprotein enzymes contain **flavin mononucleotide (FMN)** or **flavin adenine dinucleotide (FAD)** as prosthetic groups. FMN and FAD are formed in the body from the vitamin **riboflavin** (Chapter 44). FMN and FAD are usually tightly—but not covalently—bound to their respective apoenzyme proteins. Metalloflavoproteins contain one or more metals as essential cofactors. Examples of flavoprotein enzymes include **L-amino acid oxidase**, an FMN-linked enzyme found in kidney with general specificity for the oxidative deamination of the naturally occurring L-amino acids; **xanthine oxidase**, which contains molybdenum and plays an important role in the conversion of purine bases to uric acid (Chapter 33), and of particular significance in uricotelic animals (Chapter 28); and **aldehyde dehydrogenase**, an FAD-linked enzyme present in mammalian livers, which contains molybdenum and nonheme iron and acts upon aldehydes and N-heterocyclic substrates. The mechanisms of oxidation and reduction of these enzymes are complex. Evidence suggests a two-step reaction as shown in Figure 12-2.

**Figure 12-2**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oxidoreduction of isoalloxazine ring in flavin nucleotides via a semiquinone (free radical) intermediate (center).

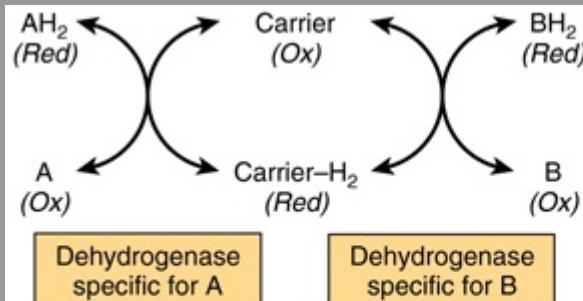
\*The term "oxidase" is sometimes used collectively to denote all enzymes that catalyze reactions involving molecular oxygen.

## DEHYDROGENASES CANNOT USE OXYGEN AS A HYDROGEN ACCEPTOR

There are a large number of enzymes in the dehydrogenase class. They perform the following two main functions:

- Transfer of hydrogen from one substrate to another in a coupled oxidation-reduction reaction (**Figure 12–3**). These dehydrogenases are specific for their substrates but often utilize common coenzymes or hydrogen carriers, eg, NAD<sup>+</sup>. Since the reactions are reversible, these properties enable reducing equivalents to be freely transferred within the cell. This type of reaction, which enables one substrate to be oxidized at the expense of another, is particularly useful in enabling oxidative processes to occur in the absence of oxygen, such as during the anaerobic phase of glycolysis (**Figure 18–2**).
- Transfer of electrons in the **respiratory chain** of electron transport from substrate to oxygen (**Figure 13–3**).

**Figure 12–3**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

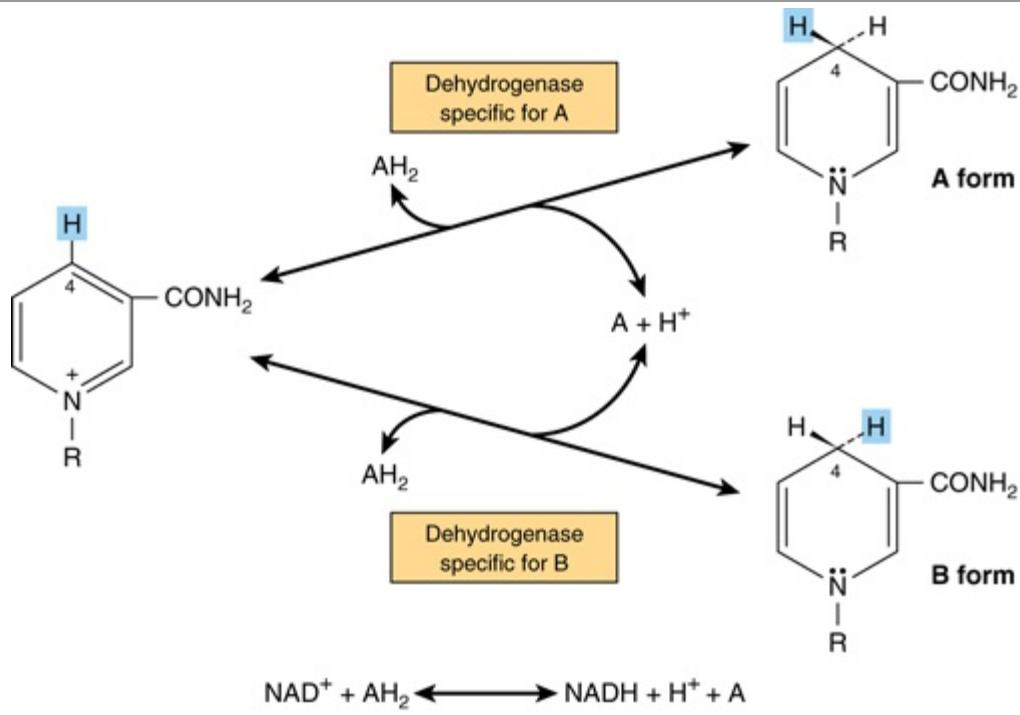
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oxidation of a metabolite catalyzed by coupled dehydrogenases.

## Many Dehydrogenases Depend on Nicotinamide Coenzymes

These dehydrogenases use **nicotinamide adenine dinucleotide (NAD<sup>+</sup>)** or **nicotinamide adenine dinucleotide phosphate (NADP<sup>+</sup>)**—or both—which are formed in the body from the vitamin **niacin** (Chapter 44). The coenzymes are reduced by the specific substrate of the dehydrogenase and reoxidized by a suitable electron acceptor (**Figure 12–4**). They are able to freely and reversibly dissociate from their respective apoenzymes.

**Figure 12–4**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Mechanism of oxidation and reduction of nicotinamide coenzymes.** There is stereospecificity about position 4 of nicotinamide when it is reduced by a substrate  $\text{AH}_2$ . One of the hydrogen atoms is removed from the substrate as a hydrogen nucleus with two electrons (hydride ion,  $\text{H}^-$ ) and is transferred to the 4 position, where it may be attached in either the A or the B form according to the specificity determined by the particular dehydrogenase catalyzing the reaction. The remaining hydrogen of the hydrogen pair removed from the substrate remains free as a hydrogen ion.

Generally, **NAD-linked dehydrogenases** catalyze oxidoreduction reactions in the oxidative pathways of metabolism, particularly in glycolysis (Chapter 18), in the citric acid cycle (Chapter 17), and in the respiratory chain of mitochondria (Chapter 13). NADP-linked dehydrogenases are found characteristically in reductive syntheses, as in the extramitochondrial pathway of fatty acid synthesis (Chapter 23) and steroid synthesis (Chapter 26)—and also in the pentose phosphate pathway (Chapter 21).

## Other Dehydrogenases Depend on Riboflavin

The **flavin groups** associated with these dehydrogenases are similar to FMN and FAD occurring in oxidases. They are generally more tightly bound to their apoenzymes than are the nicotinamide coenzymes. Most of the **riboflavin-linked dehydrogenases** are concerned with electron transport in (or to) the respiratory chain (Chapter 13). **NADH dehydrogenase** acts as a carrier of electrons between NADH and the components of higher redox potential (**Figure 13–3**). Other dehydrogenases such as **succinate dehydrogenase**, **acyl-CoA**

**dehydrogenase**, and **mitochondrial glycerol-3-phosphate dehydrogenase** transfer reducing equivalents directly from the substrate to the respiratory chain (**Figure 13–5**). Another role of the flavin-dependent dehydrogenases is in the dehydrogenation (by **dihydrolipoyl dehydrogenase**) of reduced lipoate, an intermediate in the oxidative decarboxylation of pyruvate and  $\alpha$ -ketoglutarate (**Figures 13–5 & 18–5**). The **electron-transferring flavoprotein (ETF)** is an intermediary carrier of electrons between acyl-CoA dehydrogenase and the respiratory chain (**Figure 13–5**).

## Cytochromes May Also Be Regarded as Dehydrogenases

The **cytochromes** are iron-containing hemoproteins in which the iron atom oscillates between  $\text{Fe}^{3+}$  and  $\text{Fe}^{2+}$  during oxidation and reduction. Except for cytochrome oxidase (previously described), they are classified as dehydrogenases. In the respiratory chain, they are involved as carriers of electrons from flavoproteins on the one hand to cytochrome oxidase on the other (**Figure 13–5**). Several identifiable cytochromes occur in the respiratory chain, ie, cytochromes *b*, *c<sub>1</sub>*, *c*, and cytochrome oxidase. Cytochromes are also found in other locations, eg, the endoplasmic reticulum (cytochromes P450 and *b5*), and in plant cells, bacteria, and yeasts.

## HYDROPEROXIDASES USE HYDROGEN PEROXIDE OR AN ORGANIC PEROXIDE AS SUBSTRATE

Two type of enzymes found both in animals and plants fall into this category: **peroxidases** and **catalase**.

Hydroperoxidases protect the body against harmful peroxides. Accumulation of peroxides can lead to generation of free radicals, which in turn can disrupt membranes and perhaps cause diseases including cancer and atherosclerosis (see Chapters 15 and 44).

### Peroxidases Reduce Peroxides Using Various Electron Acceptors

Peroxidases are found in milk and in leukocytes, platelets, and other tissues involved in eicosanoid metabolism (Chapter 23).

The prosthetic group is protoheme. In the reaction catalyzed by peroxidase, hydrogen peroxide is reduced at the expense of several substances that will act as electron acceptors, such as ascorbate, quinones, and cytochrome c. The reaction catalyzed by peroxidase is complex, but the overall reaction is as follows:

**PEROXIDASE**



In erythrocytes and other tissues, the enzyme **glutathione peroxidase**, containing **selenium** as a prosthetic group, catalyzes the destruction of  $\text{H}_2\text{O}_2$  and lipid hydroperoxides through the conversion of reduced glutathione to its oxidized form, protecting membrane lipids and hemoglobin against oxidation by peroxides (Chapter 21).

### Catalase Uses Hydrogen Peroxide as Electron Donor & Electron Acceptor

**Catalase** is a hemoprotein containing four heme groups. In addition to possessing peroxidase activity, it is able to use one molecule of  $\text{H}_2\text{O}_2$  as a substrate electron donor and another molecule of  $\text{H}_2\text{O}_2$  as an oxidant or electron acceptor.

**CATALASE**



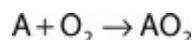
Under most conditions in vivo, the peroxidase activity of catalase seems to be favored. Catalase is found in blood, bone marrow, mucous membranes, kidney, and liver. It functions to destroy hydrogen peroxide formed by the action of oxidases. **Peroxisomes** are found in many tissues, including liver. They are rich in oxidases and in catalase. Thus, the enzymes that produce  $\text{H}_2\text{O}_2$  are grouped with the enzyme that breaks it down. However, mitochondrial and microsomal electron transport systems as well as xanthine oxidase must be considered as additional sources of  $\text{H}_2\text{O}_2$ .

## OXYGENASES CATALYZE THE DIRECT TRANSFER & INCORPORATION OF OXYGEN INTO A SUBSTRATE MOLECULE

Oxygenases are concerned with the synthesis or degradation of many different types of metabolites. They catalyze the incorporation of oxygen into a substrate molecule in two steps: (1) oxygen is bound to the enzyme at the active site and (2) the bound oxygen is reduced or transferred to the substrate. Oxygenases may be divided into two subgroups, dioxygenases and monooxygenases.

### Dioxygenases Incorporate Both Atoms of Molecular Oxygen into the Substrate

The basic reaction catalyzed by dioxygenases is shown below:



Examples include the liver enzymes, **homogentisate dioxygenase** (oxidase) and **3-hydroxyanthranilate dioxygenase** (oxidase), which contain iron; and **L-tryptophan dioxygenase** (tryptophan pyrolase) (Chapter 29), which utilizes heme.

### Monoxygenases (Mixed-Function Oxidases, Hydroxylases) Incorporate Only One Atom of Molecular Oxygen into the Substrate

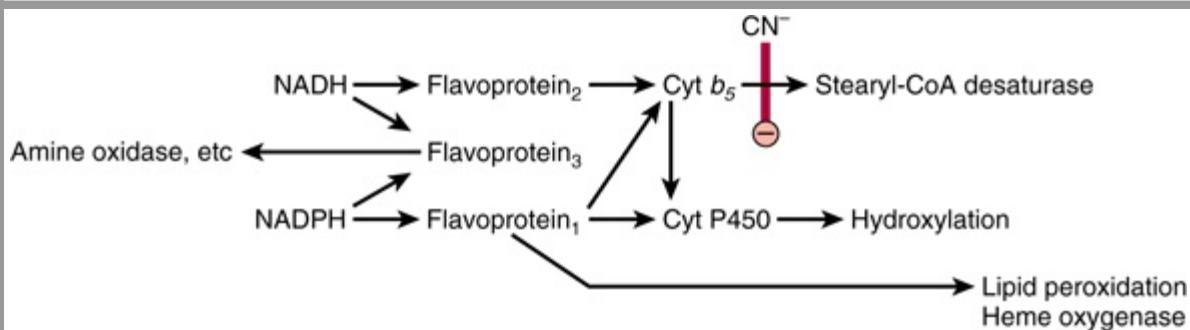
The other oxygen atom is reduced to water, an additional electron donor or cosubstrate (Z) being necessary for this purpose:



### Cytochromes P450 Are Monoxygenases Important for the Detoxification of Many Drugs & for the Hydroxylation of Steroids

**Cytochromes P450** are an important superfamily of heme-containing monooxygenases, and >50 such enzymes have been found in the human genome. These cytochromes are located mainly in the endoplasmic reticulum in the liver and intestine, but are also found in the mitochondria in some tissues. Both NADH and NADPH donate reducing equivalents for the reduction of these cytochromes (Figure 12–5), which in turn are oxidized by substrates in a series of enzymatic reactions collectively known as the **hydroxylase cycle** (Figure 12–6). In the endoplasmic reticulum of the liver, cytochromes P450 are found together with **cytochrome b<sub>5</sub>** and have a major role in drug metabolism and detoxification; they are responsible for about 75% of the modification and degradation of drugs which occurs in the body. The rate of detoxification of many medicinal drugs by cytochromes P450 determines the duration of their action. Benzpyrene, aminopyrine, aniline, morphine, and benzphetamine are hydroxylated, increasing their solubility and aiding their excretion. Many drugs such as phenobarbital have the ability to induce the synthesis of cytochromes P450.

**Figure 12–5**



Source: Murray RK, Bender DA, Botham KM, Kennelly PJ, Rodwell VW, Weil PA: *Harper's Illustrated Biochemistry*, 29th Edition: [www.accessmedicine.com](http://www.accessmedicine.com)

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

**Electron transport chain in the endoplasmic reticulum.** Cyanide (CN<sup>-</sup>) inhibits the indicated step.

**Figure 12–6**