Dedication

TO OUR TEACHERS AND OUR STUDENTS

About the authors

JEREMY M. BERG has been Professor and Director (Department Chairperson) of Biophysics and Biophysical Chemistry at Johns Hopkins University School of Medicine since 1990. He received his B.S. and M.S. degrees in Chemistry from Stanford (where he learned X-ray crystallography with Keith Hodgson and Lubert Stryer) and his Ph.D. in Chemistry from Harvard with Richard Holm. He then completed a postdoctoral fellowship with Carl Pabo. Professor Berg is recipient of the American Chemical Society Award in Pure Chemistry (1994), the Eli Lilly Award for Fundamental Research in Biological Chemistry (1995), the Maryland Outstanding Young Scientist of the Year (1995), and the Harrison Howe Award (1997). While at Johns Hopkins, he has received the W. Barry Wood Teaching Award (selected by medical students), the Graduate Student Teaching Award, and the Professor's Teaching Award for the Preclinical Sciences. He is co-author, with Stephen Lippard, of the text *Principles of Bioinorganic Chemistry*.

JOHN L. TYMOCZKO is the Towsley Professor of Biology at Carleton College, where he has taught since 1976. He currently teaches Biochemistry, Biochemistry Laboratory, Oncogenes and the Molecular Biology of Cancer, and Exercise Biochemistry and co-teaches an introductory course, Bioenergetics and Genetics. Professor Tymoczko received his B.A. from the University of Chicago in 1970 and his Ph.D. in Biochemistry from the University of Chicago with Shutsung Liao at the Ben May Institute for Cancer Research. He followed that with a post-doctoral position with Hewson Swift of the Department of Biology at the University of Chicago. Professor Tymoczko's research has focused on steroid receptors, ribonucleoprotein particles, and proteolytic processing enzymes.

LUBERT STRYER is currently Winzer Professor in the School of Medicine and Professor of Neurobiology at Stanford University, where he has been on the faculty since 1976. He received his M.D. from Harvard Medical School. Professor Stryer has received many awards for his research, including the Eli Lilly Award for Fundamental Research in Biological Chemistry (1970) and the Distinguished Inventors Award of the Intellectual Property Owners' Association. He was elected to the National Academy of Sciences in 1984. Professor Stryer was formerly the President and Scientific Director of the Affymax Research Institute. He is a founder and a member of the Scientific Advisory Board of Senomyx, a company that is using biochemical knowledge to develop new and improved flavor and fragrance molecules for use in consumer products. The publication of the first edition of his text *Biochemistry* in 1975 transformed the teaching of biochemistry.

Preface

For more than 25 years, and through four editions, Stryer's *Biochemistry* has laid out this beautiful subject in an exceptionally appealing and lucid manner. The engaging writing style and attractive design have made the text a pleasure for our students to read and study throughout our years of teaching. Thus, we were delighted to be given the opportunity to participate in the revision of this book. The task has been exciting and somewhat daunting, doubly so because of the dramatic changes that are transforming the field of biochemistry as we move into the twenty-first century. Biochemistry is rapidly progressing from a science performed almost entirely at the laboratory bench to one that may be explored through computers. The recently developed ability to determine entire genomic sequences has provided the data needed to accomplish massive comparisons of derived protein sequences, the results of which may be used to formulate and test hypotheses about biochemical function. The power of these new methods is explained by the impact of evolution: many molecules and biochemical pathways have been generated by duplicating and modifying existing ones. Our challenge in writing the fifth edition of *Biochemistry* has been to introduce this philosophical shift in biochemistry while maintaining the clear and inviting style that has distinguished the preceding four editions.Figure 9.44

A New Molecular Evolutionary Perspective

How should these evolution-based insights affect the teaching of biochemistry? Often macromolecules with a common evolutionary origin play diverse biological roles yet have many structural and mechanistic features in common. An example is a protein family containing macromolecules that are crucial to moving muscle, to transmitting the information that adrenaline is present in the bloodstream, and to driving the formation of chains of amino acids. The key features of such a protein family, presented to the student once in detail, become a model that the student can apply each time that a new member of the family is encountered. The student is then able to focus on how these features, observed in a new context, have been adapted to support other biochemical processes. Throughout the text, a stylized tree icon is  positioned at the start of discussions focused primarily on protein homologies and evolutionary origins.

Two New Chapters.

To enable students to grasp the power of these insights, two completely new chapters have been added. The first, "Biochemical Evolution" (Chapter 2), is a brief tour from the origin of life to the development of multicellular organisms. On one level, this chapter provides an introduction to biochemical molecules and pathways and their cellular context. On another level, it attempts to deepen student understanding by examining how these molecules and pathways arose in response to key biological challenges. In addition, the evolutionary perspective of Chapter 2 makes some apparently peculiar aspects of biochemistry more reasonable to students. For example, the presence of ribonucleotide fragments in biochemical cofactors can be accounted for by the likely occurrence of an early world based largely on RNA. The second new chapter, "Exploring Evolution" (Chapter 7), develops the conceptual basis for the comparison of protein and nucleic acid sequences. This chapter parallels "Exploring Proteins" (Chapter 4) and "Exploring Genes" (Chapter 6), which have thoughtfully examined experimental techniques in earlier editions. Its goal is to enable students to use the vast information available in sequence and structural databases in a critical and effective manner.

Organization of the Text.

The evolutionary approach influences the organization of the text, which is divided into four major parts. As it did in the preceding edition, Part I introduces the language of biochemistry and the structures of the most important classes of biological molecules. The remaining three parts correspond to three major evolutionary challenges—namely, the interconversion of different forms of energy, molecular reproduction, and the adaptation of cells and organisms to changing environments. This arrangement parallels the evolutionary path outlined in Chapter 2 and naturally flows from the simple to the more complex.

PART I, the molecular design of life, introduces the most important classes of biological macromolecules, including proteins, nucleic acids, carbohydrates, and lipids, and presents the basic concepts of catalysis and enzyme action. Here are two examples of how an evolutionary perspective has shaped the material in these chapters:

- **Chapter 9 ,** on catalytic strategies, examines four classes of enzymes that have evolved to meet specific challenges: promoting a fundamentally slow chemical reaction, maximizing the absolute rate of a reaction, catalyzing a reaction at one site but not at many alternative sites, and preventing a deleterious side reaction. In each case, the text considers the role of evolution in fine-tuning the key property.
- **Chapter 13 ,** on membrane channels and pumps, includes the first detailed three-dimensional structures of an ion channel and an ion pump. Because most other important channels and pumps are evolutionarily related to these proteins, these two structures provide powerful frameworks for examining the molecular basis of the action of these classes of molecules, so important for the functioning of the nervous and other systems.

PART II, transducing and storing energy, examines pathways for the interconversion of different forms of energy. Chapter 15, on signal transduction, looks at how DNA fragments encoding relatively simple protein modules, rather than entire proteins, have been mixed and matched in the course of evolution to generate the wiring that defines signal-transduction pathways. The bulk of Part II discusses pathways for the generation of

ATP and other energy-storing molecules. These pathways have been organized into groups that share common enzymes. The component reactions can be examined once and their use in different biological contexts illustrated while these reactions are fresh in the students' minds.

- Chapter 16 covers both glycolysis and gluconeogenesis. These pathways are, in some ways, the reverse of each other, and a core of enzymes common to both pathways catalyze many of the steps in the center of the pathways. Covering the pathways together makes it easy to illustrate how free energy enters to drive the overall process either in the direction of glucose degradation or in the direction of glucose synthesis.
- Chapter 17, on the citric acid cycle, ties together through evolutionary insights the pyruvate dehydrogenase complex, which feeds molecules into the citric acid cycle, and the α-ketoglutarate dehydrogenase complex, which catalyzes one of the key steps in the cycle itself.Figure 15.34
- Oxidative phosphorylation, in **Chapter 18** , is immediately followed in **Chapter 19** by the light reactions of photosynthesis to emphasize the many common chemical features of these pathways.
- The discussion of the light reactions of photosynthesis in **Chapter 19** leads naturally into a discussion of the dark reactions—that is, the components of the Calvin cycle—in **Chapter 20** . This pathway is naturally linked to the pentose phosphate pathway, also covered in **Chapter 20** , because in both pathways common enzymes interconvert three-, four-, five-, six-, and seven-carbon sugars.

PART III, synthesizing the molecules of life, focuses on the synthesis of biological macromolecules and their components.

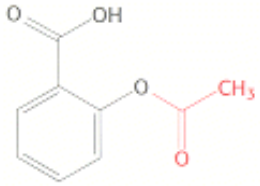
- Chapter 24, on the biosynthesis of amino acids, is linked to the preceding chapter on amino acid degradation by a family of enzymes that transfer amino groups to and from the carbon frameworks of amino acids.
- Chapter 25 covers the biosynthesis of nucleotides, including the role of amino acids as biosynthetic precursors. A key evolutionary insight emphasized here is that many of the enzymes in these pathways are members of the same family and catalyze analogous chemical reactions. The focus on enzymes and reactions common to these biosynthetic pathways allows students to understand the logic of the pathways, rather than having to memorize a set of seemingly unrelated reactions.
- Chapters 27, 28, and 29 cover DNA replication, recombination, and repair; RNA synthesis and splicing; and protein synthesis. Evolutionary connections between prokaryotic systems and eukaryotic systems reveal how the basic biochemical processes have been adapted to function in more-complex biological systems. The recently elucidated structure of the ribosome gives students a glimpse into a possible early RNA world, in which nucleic acids, rather than proteins, played almost all the major roles in catalyzing important pathways.

PART IV, responding to environmental changes, looks at how cells sense and adapt to changes in their environments. Part IV examines, in turn, sensory systems, the immune system, and molecular motors and the cytoskeleton. These chapters illustrate how signaling and response processes, introduced earlier in the text, are integrated in multicellular organisms to generate powerful biochemical systems for detecting and responding to environmental changes. Again, the adaptation of proteins to new roles is key to these discussions.

Integrated Chemical Concepts

We have attempted to integrate chemical concepts throughout the text. They include the mechanistic basis for the action of selected enzymes, the thermodynamic basis for the folding and assembly of proteins and other macromolecules, and the structures and chemical reactivity of the common cofactors. These fundamental topics underlie our understanding of all biological processes. Our goal is not to provide an encyclopedic examination of enzyme reaction mechanisms. Instead, we have selected for examination at a more detailed chemical level specific topics that will enable students to understand how the chemical features help meet the biological needs.

Chemical insight often depends on a clear understanding of the structures of biochemical molecules. We have taken considerable care in preparing stereochemically accurate depictions of these molecules where appropriate. These structures should make it easier for the student to develop an intuitive feel for the shapes of molecules and comprehension of how these shapes affect reactivity.



Aspirin (acetylsalicyclic acid)

Newly Updated to Include Recent Discoveries

Given the breathtaking pace of modern biochemistry, it is not surprising that there have been major developments since the publication of the fourth edition. Foremost among them is the sequencing of the human genome and the genomes of many simpler organisms. The text's evolutionary framework allows us to naturally incorporate information from these historic efforts. The determination of the three-dimensional structures of proteins and macromolecular assemblies also has been occurring at an astounding pace.

- As noted earlier, the discussion of excitable membranes in **Chapter 13** incorporates the detailed structures of an ion channel (the prokaryotic potassium channel) and an ion pump (the sacroplasmic reticulum calcium ATPase). Figure 9.21
- Great excitement has been generated in the signal transduction field by the first determination of the structure of a seven-transmembrane-helix receptor—the visual system protein rhodopsin—discussed in **Chapters 15** and **32**
- The ability to describe the processes of oxidative phosphorylation in **Chapter 18** has been greatly aided by the determination of the structures for two large membrane protein complexes: cytochrome *c* oxidase and cytochrome bc_1.
- Recent discoveries regarding the three-dimensional structure of ATP synthase are covered in **Chapter 18** , including the remarkable fact that parts of the enzyme rotate in the course of catalysis.
- The determination of the structure of the ribosome transforms the discussion of protein synthesis in **Chapter 29** .
- The elucidation of the structure of the nucleosome core particle—a large protein–DNA complex— facilitates the description in **Chapter 31** of key processes in eukaryotic gene regulation.

Finally, each of the three chapters in Part IV is based on recent structural conquests.

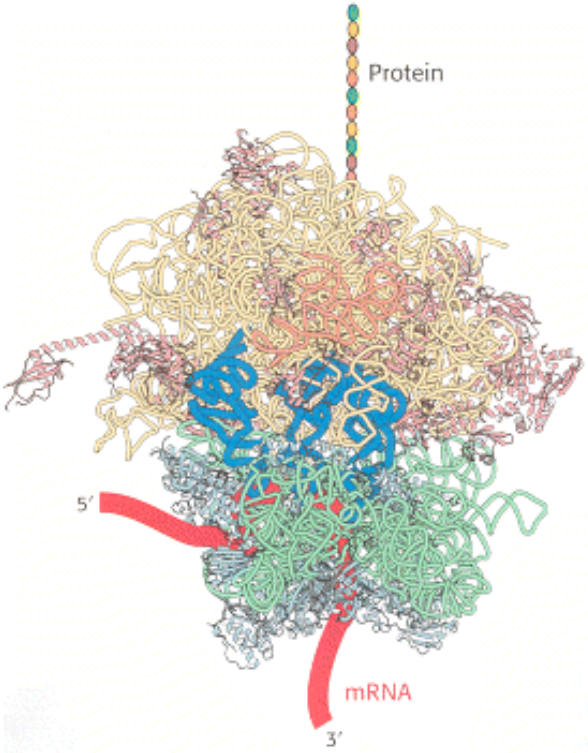
- The ability to grasp key concepts in sensory systems (**Chapter 32**) is aided by the structures of rhodopsin and the aforementioned ion channel.
- **Chapter 33** , on the immune system, now includes the more recently determined structure of the T-cell receptor and its complexes.
- The determination of the structures of the molecular motor proteins myosin and kinesin first revealed the evolutionary connections on which **Chapter 34** , on molecular motors, is based.

New and Improved Illustrations

The relation of structure and function has always been a dominant theme of *Biochemistry*. This relation becomes even clearer to students using the fifth edition through the extensive use of molecular models. These models are superior to those in the fourth edition in several ways.

- All have been designed and rendered by one of us (JMB), with the use of MOLSCRIPT, to emphasize the most important structural features. The philosophy of the authors is that the reader should be able to write the caption from looking at the picture.
- We have chosen ribbon diagrams as the most effective, clearest method of conveying molecular structure. All molecular diagrams are rendered in a consistent style. Thus students are able to compare structures easily and to develop familiarity and facility in interpreting the models. Labels highlight key features of the molecular models.
- Many new molecular models have been added, serving as sources of structural insight into additional molecules and in some cases affording multiple views of the same molecule.

In addition to the molecular models, the fifth edition includes more diagrams providing an overview of pathways and processes and setting processes in their biological context.



New Pedagogical Features

The fifth edition of *Biochemistry* supplies additional tools to assist students in learning the subject matter.

Icons.

Icons are used to highlight three categories of material, making these topics easier to locate for the interested student or teacher.

A *caduceus* signals the beginning of a clinical application.

A *stylized tree* marks sections or paragraphs that primarily or exclusively explore evolutionary aspects of biochemistry.

A *mouse and finger* point to references to animations on the text's Web site (www.whfreeman.com/biochem5) for those students who wish to reinforce their understanding of concepts by using the electronic media.

More Problems.

The number of problems has increased by 50%. Four new categories of problem have been created to develop specific skills.

Mechanism problems ask students to suggest or elaborate a chemical mechanism.
Data interpretation problems ask questions about a set of data provided in tabulated or graphic form. These exercises give students a sense of how scientific conclusions are reached.

Chapter integration problems require students to use information from multiple chapters to reach a solution. These problems reinforce awareness of the interconnectedness of the different aspects of biochemistry.
Media problems encourage and assist students in taking advantage of the animations and tutorials provided on our Web site. Media problems are found both in the book and on our Web site.Figure 15.23

New Chapter Outline and Key Terms.

An outline at the beginning of each chapter gives major headings and serves as a framework for students to use in organizing the information in the chapter. The major headings appear again in the chapter's summary, again helping to organize information for easier review. A set of key terms also helps students focus on and review the important concepts.Figure 17.4

Tools and Techniques

The fifth edition of *Biochemistry* offers three chapters that present the tools and techniques of biochemistry: "Exploring Proteins" (Chapter 4), "Exploring Genes" (Chapter 6), and "Exploring Evolution" (Chapter 7). Additional experimental techniques are presented elsewhere throughout the text, as appropriate.

Exploring Proteins (**Chapter 4**)

Protein purification Section 4.1

Differential centrifugation Section 4.1.2

Salting out Section 4.1.3

Dialysis Section 4.1.3

Gel-filtration chromatography Section 4.1.3

Ion-exchange chromatography Section 4.1.3

Affinity chromatography Section 4.1.3

High-pressure liquid chromatography Section 4.1.3

Gel electrophoresis Section 4.1.4

Isoelectric focusing Section 4.1.4

Two-dimensional electrophoresis Section 4.1.4

Qualitative and quantitative evaluation of protein purification Section 4.1.5

Ultracentrifugation Section 4.1.6

Mass spectrometry (MALDI-TOF) Section 4.1.7

Peptide mass fingerprinting Section 4.1.7

Edman degradation Section 4.2

Protein sequencing Section 4.2

Production of polyclonal antibodies Section 4.3.1

Production of monoclonal antibodies Section 4.3.2

Enzyme-linked immunosorbent assay (ELISA) Section 4.3.3

Western blotting Section 4.3.4

Fluorescence microscopy Section 4.3.5

Green fluorescent protein as a marker Section 4.3.5

Immunoelectron microscopy Section 4.3.5

Automated solid-phase peptide synthesis Section 4.4

Nuclear magnetic resonance spectroscopy Section 4.5.1

NOESY spectroscopy Section 4.5.1

X-ray crystallography Section 4.5.2

Exploring Proteins (other chapters)

Basis of fluorescence in green fluorescent protein Section 3.6.5

Time-resolved crystallography Section 8.3.2

Using fluorescence spectroscopy to analyze enzyme– substrate interactions Section 8.3.2

Using irreversible inhibitors to map the active site Section 8.5.2

Using transition state analogs to study enzyme active sites Section 8.5.3

Catalytic antibodies as enzymes Section 8.5.4

Exploring Genes (**Chapter 6**)

Restriction-enzyme analysis Sections 6.1.1 and 6.1.2

Southern and Northern blotting techniques Section 6.1.2

Sanger dideoxy method of DNA sequencing Section 6.1.3

Solid-phase analysis of nucleic acids Section 6.1.4

Polymerase chain reaction (PCR) Section 6.1.5

Recombinant DNA technology Sections 6.2-6.4

DNA cloning in bacteria Sections 6.2.2 and 6.2.3

Chromosome walking Section 6.2.4

Cloning of eukaryotic genes in bacteria Section 6.3.1

Examining expression levels (gene chips) Section 6.3.2

Introducing genes into eukaryotes Section 6.3.3

Transgenic animals Section 6.3.4

Gene disruption Section 6.3.5

Tumor-inducing plasmids Section 6.3.6

Site-specific mutagenesis Section 6.4

Exploring Genes (other chapters)

Density-gradient equilibrium sedimentation Section 5.2.2

Footprinting technique for isolating and characterizing promoter sites Section 28.1.1

Chromatin immunoprecipitation (ChIP) Section 31.2.3

Exploring Evolution (**Chapter 7**)

Sequence-comparison methods Section 7.2

Sequence-alignment methods Section 7.2

Estimating the statistical significance of alignments (by shuffling) Section 7.2.1

Substitution matrices Section 7.2.2

Sequence templates Section 7.3.2

Self-diagonal plots for finding repeated motifs Section 7.3.3

Mapping secondary structures through RNA sequence comparisons Section 7.3.5

Construction of evolutionary trees Section 7.4

Combinatorial chemistry Section 7.5.2

Other Techniques

Sequencing of carbohydrates by using MALDI-TOF mass spectrometry Section 11.3.7

Use of liposomes to investigate membrane permeability Section 12.4.1

Use of hydropathy plots to locate transmembrane helices Section 12.5.4

Fluorescence recovery after photobleaching (FRAP) for measuring lateral diffusion in membranes Section 12.6

Patch-clamp technique for measuring channel activity Section 13.5.1

Measurement of redox potential Section 18.2.1

Functional magnetic resonance imaging (fMRI) Section 32.1.3

Animated Techniques: Animated explanations of experimental techniques used for exploring genes and proteins are available at **www.whfreeman.com/biochem5**

Clinical Applications

This icon signals the start of a clinical application in the text. Additional, briefer clinical correlations appear without the icon in the text as appropriate.

Prion diseases Section 3.6.1

Scurvy and collagen stabilization Section 3.6.5

Antigen detection with ELISA Section 4.3.3

Vasopressin deficiency Section 4.4

Action of penicillin Section 8.5.5

Water-soluble vitamins Section 8.6.1

Fat-soluble vitamins in blood clotting and vision Section 8.6.2

Protease inhibitors Section 9.1.7

Carbonic anhydrase and osteopetrosis Section 9.2

Use of isozymes to diagnose tissue damage Section 10.3

Steatorrhea in liver disease Section 22.1.1

Carnitine deficiency Section 22.2.3

Zellweger syndrome Section 22.3.4

Diabetic ketosis Section 22.3.6

Use of fatty acid synthase inhibitors as drugs Section 22.4.9

Effects of aspirin on signaling pathways Section 22.6.2

Cervical cancer and ubiquitin Section 23.2.1

Protein degradation and the immune response Section 23.2.3

Inherited defects of the urea cycle (hyperammonemia) Section 23.4.4

Inborn errors of amino acid degradation Section 23.6

High homocysteine levels and vascular disease Section 24.2.9

Inherited disorders of porphyrin metabolism Section 24.4.4

Anticancer drugs that block the synthesis of thymidylate Section 25.3.3

Pellagra Section 25.5

Gout Section 25.6.1

Lesch-Nyhan syndrome Section 25.6.2

Disruption of lipid metabolism as the cause of respiratory distress syndrome and Tay-Sachs disease Section 26.1.6

Diagnostic use of blood cholesterol levels Section 26.3.2

Hypercholesteremia and atherosclerosis Section 26.3.5

Clinical management of cholesterol levels Section 26.3.6

Rickets and vitamin D Section 26.4.7

Antibiotics that target DNA gyrase Section 27.3.4

Defective repair of DNA and cancer Section 27.6.5

Huntington chorea Section 27.6.6

Detection of carcinogens (Ames test) Section 27.6.7

Molecular Evolution

This icon signals the start of many discussions that highlight protein commonalities or other molecular evolutionary insights that provide a framework to help students organize information.

Why is glucose a prominent fuel? Section 16.0.1

A common binding site in dehydrogenases Section 16.1.10

The major facilitator (MF) superfamily of transporters Section 16.2.4

Isozymic forms of lactate dehydrogenase Section 16.4.2

Evolutionary relationship of glycolysis and gluconeogenesis Section 16.4.3

Decarboxylation of α-ketoglutarate and pyruvate Section 17.1.6

Evolution of succinyl CoA synthetase Section 17.1.7

Evolutionary history of the citric acid cycle Section 17.3.3

Endosymbiotic origins of mitochondria Section 18.1.2

Conservation of cytochrome c structure Section 18.3.7

Common features of ATP synthase and G proteins Section 18.4.5

Related uncoupling proteins Section 18.6.4

Evolution of chloroplasts Section 19.1.2

Evolutionary origins of photosynthesis Section 19.6

Evolution of the C_4 pathway Section 20.2.3

Increasing sophistication of glycogen phosphorylase regulation Section 21.3.3

The α-amylase family Section 21.4.3

A recurring motif in the activation of carboxyl groups Section 22.2.2

Polyketide and nonribosomal peptide synthetases resemble fatty acid synthase Section 22.4.10

Prokaryotic counterparts of the ubiquitin pathway and the proteasome Section 23.2.4

A family of pyridoxal-dependent enzymes Section 23.3.3

Evolution of the urea cycle Section 23.4.3

The P-loop NTPase domain in nitrogenase Section 24.1.1

Recurring steps in purine ring synthesis Section 25.2.3

Ribonucleotide reductases Section 25.3

Increase in urate levels during primate evolution Section 25.6.1

The cytochrome P450 superfamily Section 26.4.3

DNA polymerases Section 27.2.1

Helicases Section 27.2.5

Evolutionary relationship of recombinases and topoisomerases Section 27.5.2

Similarities in transcriptional machinery between archaea and eukaryotes Section 28.2.4

Evolution of spliceosome-catalyzed splicing Section 28.2.4

Classes of aminoacyl-tRNA synthetases Section 29.2.5

Composition of the primordal ribosome Section 29.3.1

Evolution of molecular mimics Section 29.4.4

A family of proteins with common ligand-binding domains Section 31.1.4

Independent evolution of DNA-binding sites of regulatory proteins Section 31.1.5

CpG islands Section 31.2.5

Iron response elements Section 31.4.2

The odorant receptor family Section 32.1.1

Evolution of taste receptor mRNA Section 32.2.5

Photoreceptor evolution Section 32.3.4

The immunoglobulin fold Section 33.2

Relationship of actin to hexokinase and other prokaryotic proteins Section 34.2.2

Tubulins in the P-loop NTPase family Section 34.3.1

Supplements Supporting *Biochemistry*, Fifth Edition

The fifth edition of *Biochemistry* offers a wide selection of high-quality supplements to assist students and instructors.

For the Instructor

Print and Computerized Test Banks NEW

Marilee Benore Parsons, University of Michigan-Dearborn Print Test Bank 0-7167-4384-1; Computerized Test Bank CD-ROM (Windows/Macintosh hybrid) 0-7167-4386-8

The test bank offers more than 1700 questions posed in multiple choice, matching, and short-answer formats. The electronic version of the test bank allows instructors to easily edit and rearrange the questions or add their own material.

Instructor's Resource CD-ROM NEW

© W. H. Freeman and Company and Sumanas, Inc. 0-7167-4385-X

The Instructor's Resource CD-ROM contains all the illustrations from the text. An easy-to-use presentation manager application, Presentation Manager Pro, is provided. Each image is stored in a variety of formats and resolutions, from simple jpg and gif files to preformatted PowerPoint® slides, for instructors using other presentation programs.

Overhead Transparencies

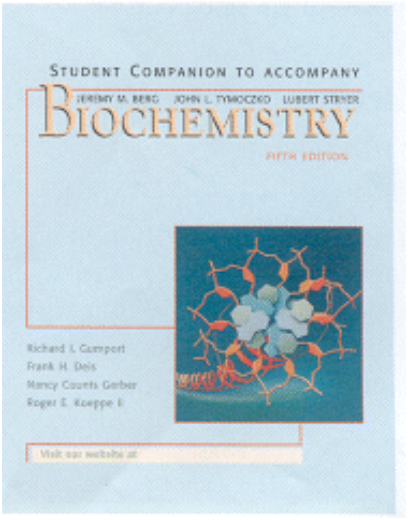
0-7167-4422-8

Full-color illustrations from the text, optimized for classroom projection, in one volume.

For the Student

Student Companion

Richard I. Gumport, College of Medicine at Urbana-Champaign, University of Illinois; Frank H. Deis, Rutgers University; and Nancy Counts Gerber, San Fransisco State University. Expanded solutions to text problems provided by Roger E. Koeppe II, University of Arkansas 0-7167-4383-3

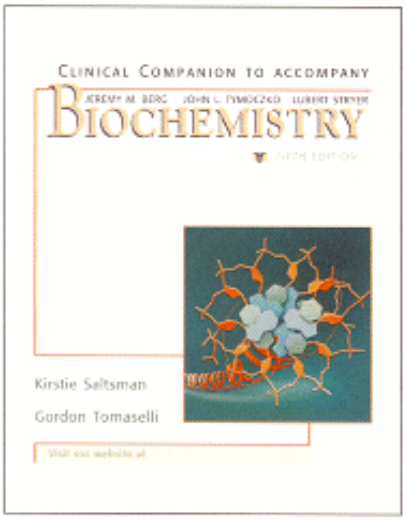


More than just a study guide, the *Student Companion* is an essential learning resource designed to meet the needs of students at all levels. Each chapter starts with a summarized abstract of the related textbook chapter. A comprehensive list of learning objectives allows students to quickly review the key concepts. A self-test feature allows students to quickly refresh their understanding, and a set of additional problems requires students to apply their knowledge of biochemistry. The complete solution to every problem in the text is provided to help students better comprehend the core ideas. Individual chapters of the *Student Companion* can be purchased and downloaded from

www.whfreeman.com/biochem5

Clinical Companion NEW

Kirstie Saltsman, Ph.D., Jeremy M. Berg, M.D., and Gordon Tomaselli, M.D., Johns Hopkins University School of Medicine 0-7167-4738-3



Designed for students and instructors interested in clinical applications, the *Clinical Companion* is a rich compendium of medical case studies and clinical discussions. It contains numerous problems and references to the textbook. Such topics as glaucoma, cystic fibrosis, Tay-Sachs disease, and autoimmune diseases are covered from a biochemical perspective.

Lecture Notebook NEW

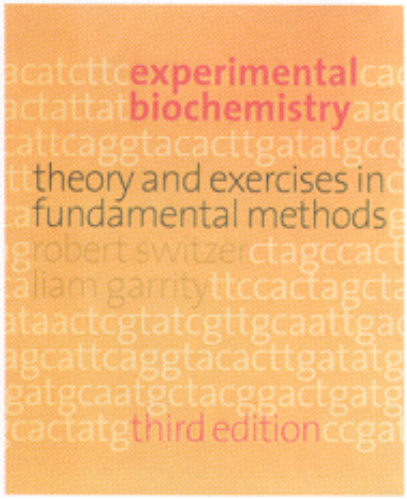
0-7167-4682-4

For students who find that they are too busy writing notes to pay attention in class, the *Lecture Notebook* brings together a black-and-white collection of illustrations from the text, arranged in the order of their appearance in the textbook, with plenty of room alongside for students to take notes.

Experimental Biochemistry, Third Edition

Robert L. Switzer, University of Illinois, and Liam F. Garrity, Pierce Chemical Corporation 0-7167-3300-5

The new edition of *Experimental Biochemistry* has been completely revised and updated to make it a perfect fit for today's laboratory course in biochemistry. It provides comprehensive coverage of important techniques used in contemporary biochemical research and gives students the background theory that they need to understand the experiments. Thoroughly classroom tested, the experiments incorporate the full range of biochemical materials in an attempt to simulate work in a research laboratory. In addition, a comprehensive appendix provides detailed procedures for preparation of reagents and materials, as well as helpful suggestions for the instructor.



Also Available Through the W. H. Freeman Custom Publishing Program

Experimental Biochemistry is designed to meet all your biochemistry laboratory needs. Visit http://custompub.whfreeman.com to learn more about creating your own laboratory manual.

Student Media Resources

This icon links materials from the book to our Web site. See the inside front cover for a complete description of the resources available at www.whfreeman.com/ biochem5.

nature publishing group npg

W. H. Freeman and Company is proud to present a weekly collection of abstracts and a monthly featured article from the *Nature* family of journals, including *Nature*, its associated monthly titles, and the recently launched *Nature Review Journals*. Please visit www.whfreeman.com/biochem5 for more information.

Acknowledgments

There is an old adage that says that you never really learn a subject until you teach it. We now know that you learn a subject even better when you write about it. Preparing the fifth edition of *Biochemistry* has provided us with a wonderful opportunity to unite our love of biochemistry and teaching and to share our enthusiasm with students throughout the world. Nonetheless, the project has also been a daunting one because so many interesting discoveries have been made since the publication of the fourth edition. The question constantly confronted us: What biochemical knowledge is most worth having? Answering this question required attempting to master as much of the new material as possible and then deciding what to include and, even harder, what to exclude.

However, we did not start from scratch. We feel both fortunate and intimidated to be writing the fifth edition of Stryer's *Biochemistry*. Fortunate, because we had as our starting point the best biochemistry book ever produced. Intimidated, because we had as our starting point the best biochemistry book ever produced, with the challenge of improving it. To the extent that we have succeeded, we have done so because of the help of many people.

Thanks go first and foremost to our students at Johns Hopkins University and Carleton College. Not a word was written or an illustration constructed without the knowledge that bright, engaged students would immediately detect vagueness or ambiguity. One of us (JMB) especially thanks the members of the Berg lab who have cheerfully tolerated years of neglect and requests to review drafts of illustrations when they would rather have been discussing their research. Particular thanks go to Dr. Barbara Amann and Kathleen

Kolish who helped preserve some order in the midst of chaos. We also thank our colleagues at Johns Hopkins University and Carleton College who supported, advised, instructed, and simply bore with us during this arduous task. One of us (JLT) was graciously awarded a grant from Carleton College to relieve him of some of his academic tasks so that he could focus more fully on the book.

We are also grateful to our colleagues throughout the world who served as reviewers for the new edition. Their thoughtful comments, suggestions, and encouragement have been of immense help to us in maintaining the excellence of the preceding editions. These reviewers are:

Mark Alper

University of California at Berkeley

L. Mario Amzel

Johns Hopkins University

Paul Azari

Colorado State University

Ruma Banerjee

University of Nebraska

Michael Barbush

Baker University

Douglas Barrick

Johns Hopkins University

Loran L. Bieber

Michigan State University

Margaret Brosnan

University of Newfoundland

Lukas K. Buehler

University of California at San Diego

C. Allen Bush

University of Maryland, Baltimore County

Tom Cech

Howard Hughes Medical Institute

Oscar P. Chilson

Washington University

Steven Clarke

University of California at Los Angeles

Philip A. Cole

Johns Hopkins University School of Medicine

Paul A. Craig

Rochester Institute of Technology

David L. Daleke

Indiana University

David Deamer

University of California at Santa Cruz

Frank H. Deis

Rutgers University

Eric S. Eberhardt

Vassar College

Duane C. Eichler

University of San Francisco School of Medicine

Stephen H. Ellis

Auburn University

Nuran Ercal

University of Missouri at Rolla

Gregg B. Fields

Florida Atlantic University

Gregory J. Gatto Jr.

Johns Hopkins University

Nancy Counts Gerber

San Francisco State University

Claiborne Glover III

University of Georgia

E. M. Gregory

Virginia Polytechnic Institute and State University

Mark Griep

University of Nebraska at Lincoln

Hebe M. Guardiola-Diaz

Trinity College

James R. Heitz

Mississippi State University

Neville R. Kallenbach

New York University

Harold Kasinsky

University of British Columbia

Dan Kirschner

Boston College

G. Barrie Kitto

University of Texas at Austin

James F. Koerner

University of Minnesota

John Koontz

University of Tennessee

Gary R. Kunkel

Texas A&M University

David O. Lambeth

University of North Dakota

Timothy Logan

Florida State University

Douglas D. McAbee

California State University at Long Beach

William R. Marcotte Jr.

Clemson University

Alan Mellors

University of Guelph

Dudley G. Moon

Albany College of Pharmacy

Carolyn M. Teschke

University of Connecticut

Dean R. Tolan

Boston University

Gordon Tollin

University of Arizona

Jeffrey M. Voigt

Albany College of Pharmacy

M. Gerard Waters

Princeton University

Linette M. Watkins

Southwest Texas State University

Gabriele Wienhausen

University of California at San Diego

James D. Willett

George Mason University

Gail R. Willsky

State University of New York at Buffalo

Dennis Winge

University of Utah

Charles F. Yocum

University of Michigan

Working with our colleagues at W. H. Freeman and Company has been a wonderful experience. We would especially like to acknowledge the efforts of the following people. Our development editor, Susan Moran, contributed immensely to the success of this project. During this process, Susan became a committed biochemistry student. Her understanding of how the subject matter, text, and illustrations, would be perceived by students and her commitment to excellence were a true inspiration. Our project editor, Georgia Lee Hadler, managed the flow of the entire project—from manuscript to final product—sometimes with a velvet glove and other times more forcefully, but always effectively. The careful

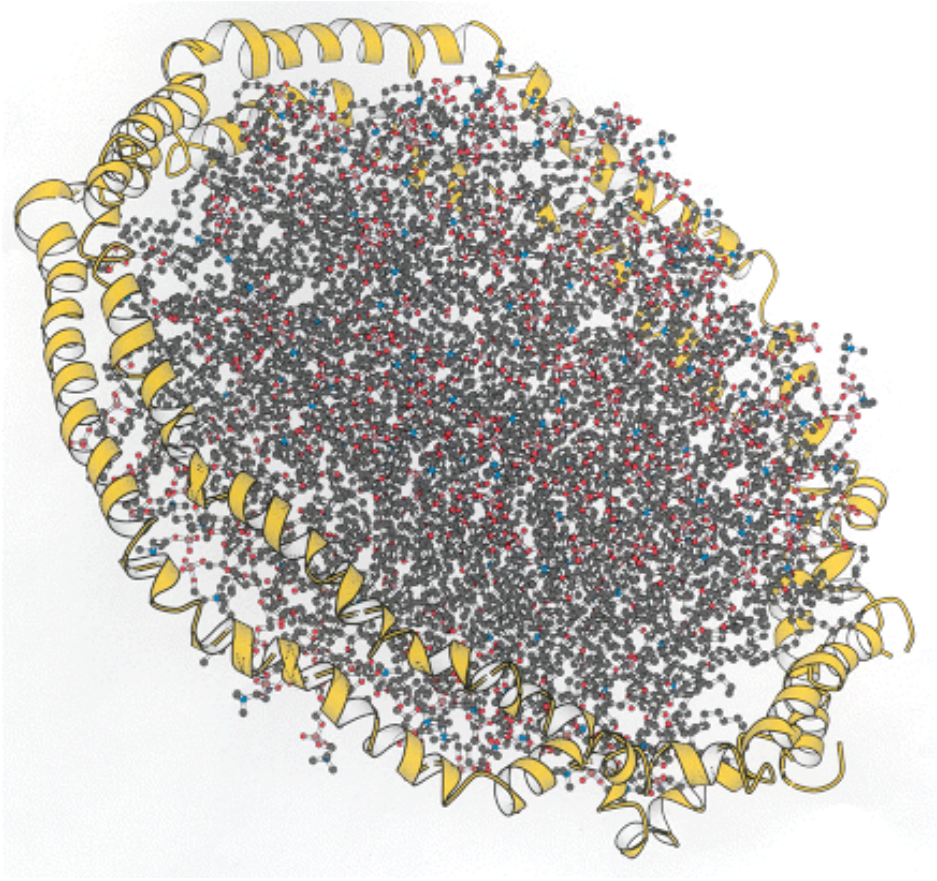
manuscript editor, Patricia Zimmerman, enhanced the text's literary consistency and clarity. Designers Vicki Tomaselli and Patricia McDermond produced a design and layout that are organizationally clear and aesthetically pleasing. The tireless search of our photo researchers, Vikii Wong and Dena Betz, for the best possible photographs has contributed effectively to the clarity and appeal of the text. Cecilia Varas, the illustration coordinator, ably oversaw the rendering of hundreds of new illustrations, and Julia DeRosa, the production manager, astutely handled all the difficulties of scheduling, composition, and manufacturing.

Neil Clarke of Johns Hopkins University, Sonia DiVittorio, and Mark Santee piloted the media projects associated with the book. Neil's skills as a teacher and his knowledge of the power and pitfalls of computers, Sonia's editing and coordination skills and her stylistic sense, and Mark's management of an ever-changing project have made the Web site a powerful supplement to the text and a lot of fun to explore. We want to acknowledge the media developers who transformed scripts into the animations you find on our Web site. For the Conceptual Insights modules we thank Nick McLeod, Koreen Wykes, Dr. Roy Tasker, Robert Bleeker, and David Hegarty, all at CADRE design. For the threedimensional molecular visualizations in the Structural Insights modules we thank Timothy Driscoll (molvisions.com—3D molecular visualization). Daniel J. Davis of the University of Arkansas at Fayetteville prepared the online quizzes.

Publisher Michelle Julet was our cheerleader, taskmaster, comforter, and cajoler. She kept us going when we were tired, frustrated, and discouraged. Along with Michelle, marketing mavens John Britch and Carol Coffey introduced us to the business of publishing. We also thank the sales people at W. H. Freeman and Company for their excellent suggestions and view of the market, especially Vice President of Sales Marie Schappert, David Kennedy, Chris Spavins, Julie Hirshman, Cindi Weiss-Goldner, Kimberly Manzi, Connaught Colbert, Michele Merlo, Sandy Manly, and Mike Krotine. We thank Elizabeth Widdicombe, President of W. H. Freeman and Company, for never losing faith in us.

Finally, the project would not have been possible without the unfailing support of our families—especially our wives, Wendie Berg and Alison Unger. Their patience, encouragement, and enthusiasm have made this endeavor possible. We also thank our children, Alex, Corey, and Monica Berg and Janina and Nicholas Tymoczko, for their forbearance and good humor and for constantly providing us a perspective on what is truly important in life.

I. The Molecular Design of Life



Part of a lipoprotein particle. A model of the structure of apolipoprotein A-I (yellow), shown surrounding sheets of lipids. The apolipoprotein is the major protein component of high-density lipoprotein particles in the blood. These particles are effective lipid transporters because the protein component provides an interface between the hydrophobic lipid chains and the aqueous environment of the bloodstream. [Based on coordinates provided by Stephen Harvey.]

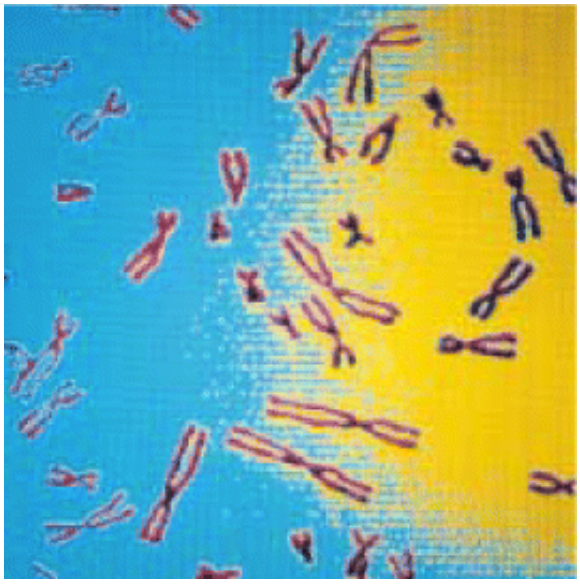
1. Prelude: Biochemistry and the Genomic Revolution

GACTTCACTTCTAATGATGATTATGGGAGAACTGGAGCCTT
CAGAGGGTAAAAATTAAGCACAGTGGAAGAATTTCATTC
TGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAG
AAAATATCTTTGGTGTTTCCTATGATGAATATAGATACAG
AAGCGTCATCAAAGCATGCCAACTAGAAGAG. . .. This string of letters A, C, G, and T is a part of a DNA sequence. Since the biochemical techniques for DNA sequencing were first developed more than three decades ago, the genomes of dozens of organisms have been sequenced, and many more such sequences will be forthcoming. The information contained in these DNA sequences promises to shed light on many fascinating and important questions. What genes in *Vibrio cholera,* the bacterium that causes cholera, for example, distinguish it from its more benign relatives? How is the development of complex organisms controlled? What are the evolutionary relationships between organisms?

Sequencing studies have led us to a tremendous landmark in the history of biology and, indeed, humanity. *A nearly complete sequence of the entire human genome* has been determined. The string of As, Cs, Gs, and Ts with which we began this book is a tiny part of the human genome sequence, which is more than 3 billion letters long. If we included the entire sequence, our opening sentence would fill more than 500,000 pages.

The implications of this knowledge cannot be overestimated. By using this blueprint for much of what it means to be

human, scientists can begin the identification and characterization of sequences that foretell the appearance of specific diseases and particular physical attributes. One consequence will be the development of better means of diagnosing and treating diseases. Ultimately, physicians will be able to devise plans for preventing or managing heart disease or cancer that take account of individual variations. Although the sequencing of the human genome is an enormous step toward a complete understanding of living systems, much work needs to be done. Where are the functional genes within the sequence, and how do they interact with one another? How is the information in genes converted into the functional characteristics of an organism? Some of our goals in the study of biochemistry are to learn the concepts, tools, and facts that will allow us to address these questions. It is indeed an exciting time, the beginning of a new era in biochemistry.



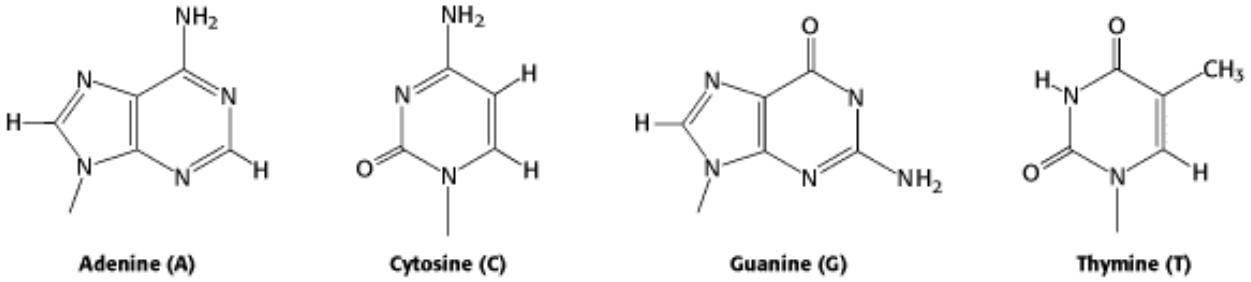
Disease and the genome. Studies of the human genome are revealing disease origins and other biochemical mysteries. Human chromosomes, left, contain the DNA molecules that constitute the human genome. The staining pattern serves to identify specific regions of a chromosome. On the right is a diagram of human chromosome 7, with band q31.2 indicated by an arrow. A gene in this region encodes a protein that, when malfunctioning, causes cystic fibrosis. [(Left) Alfred Pasieka/Peter Arnold.]

1.1. DNA Illustrates the Relation between Form and Function

The structure of DNA, an abbreviation for *d eoxyribo n ucleic a cid*, illustrates a basic principle common to all biomolecules: the intimate relation between structure and function. The remarkable properties of this chemical substance allow it to function as a very efficient and robust vehicle for storing information. We begin with an examination of the covalent structure of DNA and its extension into three dimensions.

1.1.1. DNA Is Constructed from Four Building Blocks

DNA is a *linear polymer* made up of four different monomers. It has a fixed backbone from which protrude variable substituents (Figure 1.1). The backbone is built of repeating sugar-phosphate units. The sugars are molecules of *deoxyribose* from which DNA receives its name. Joined to each deoxyribose is one of four possible bases: adenine (A), cytosine (C), guanine (G), and thymine (T).



All four bases are planar but differ significantly in other respects. Thus, the monomers of DNA consist of a sugar-phosphate unit, with one of four bases attached to the sugar. *These bases can be arranged in any order along a strand of DNA*. The order of these bases is what is displayed in the sequence that begins this chapter. For example, the first base in the sequence shown is G (guanine), the second is A (adenine), and so on. *The sequence of bases along a DNA strand constitutes the genetic information* —the instructions for assembling proteins, which themselves orchestrate the synthesis of a host of other biomolecules that form cells and ultimately organisms.

1.1.2. Two Single Strands of DNA Combine to Form a Double Helix

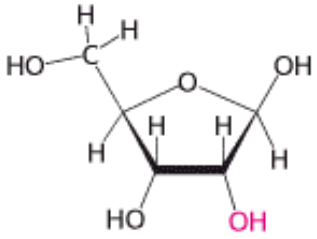
Most DNA molecules consist of not one but two strands (Figure 1.2). How are these strands positioned with respect to one another? In 1953, James Watson and Francis Crick deduced the arrangement of these strands and proposed a three-dimensional structure for DNA molecules. This structure is a *double helix* composed of two intertwined strands arranged such that the sugar-phosphate backbone lies on the outside and the bases on the inside. The key to this structure is that the bases form *specific base pairs* (bp) held together by *hydrogen bonds* (Section 1.3.1): adenine pairs with thymine (A-T) and guanine pairs with cytosine (G-C), as shown in Figure 1.3. Hydrogen bonds are much weaker than covalent bonds such as the carbon-carbon or carbon-nitrogen bonds that define the structures of the bases themselves. Such weak bonds are crucial to biochemical systems; they are weak enough to be reversibly broken in biochemical processes, yet they are strong enough, when many form simultaneously, to help stabilize specific structures such as the double helix.

The structure proposed by Watson and Crick has two properties of central importance to the role of DNA as the hereditary material. First, the structure is compatible with *any sequence of bases*. The base pairs have essentially the same shape (Figure 1.4) and thus fit equally well into the center of the double-helical structure. Second, because of base-pairing, *the sequence of bases along one strand completely determines the sequence along the other strand*. As Watson and Crick so coyly wrote: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." Thus, if the DNA double helix is separated into two single strands, each strand can act as a template for the generation of its partner strand through specific base-pair formation (Figure 1.5). *The three-dimensional structure of DNA beautifully illustrates the close connection between*

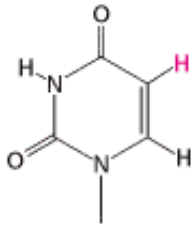
molecular form and function.

1.1.3. RNA Is an Intermediate in the Flow of Genetic Information

An important nucleic acid in addition to DNA is *r ibo n ucleic a cid (RNA)*. Some viruses use RNA as the genetic material, and even those organisms that employ DNA must first convert the genetic information into RNA for the information to be accessible or functional. Structurally, RNA is quite similar to DNA. It is a linear polymer made up of a limited number of repeating monomers, each composed of a sugar, a phosphate, and a base. The sugar is ribose instead of deoxyribose (hence, *R*NA) and one of the bases is uracil (U) instead of thymine (T). Unlike DNA, an RNA molecule usually exists as a single strand, although significant segments within an RNA molecule may be double stranded, with G pairing primarily with C and A pairing with U. This intrastrand base-pairing generates RNA molecules with complex structures and activities, including catalysis.



Ribose



Uracil (U)

RNA has three basic roles in the cell. First, it serves as the intermediate in the flow of information from DNA to protein, the primary functional molecules of the cell. The DNA is copied, or *transcribed*, into messenger RNA (mRNA), and the mRNA is *translated* into protein. Second, RNA molecules serve as adaptors that translate the information in the nucleic acid sequence of mRNA into information designating the sequence of constituents that make up a protein. Finally, RNA molecules are important functional components of the molecular machinery, called ribosomes, that carries out the translation process. As will be discussed in Chapter 2, the unique position of RNA between the storage of genetic information in DNA and the functional expression of this information as protein as well as its potential to combine genetic and catalytic capabilities are indications that RNA played an important role in the evolution of life.

1.1.4. Proteins, Encoded by Nucleic Acids, Perform Most Cell Functions

A major role for many sequences of DNA is to encode the sequences of *proteins*, the workhorses within cells, participating in essentially all processes. Some proteins are key structural components, whereas others are specific catalysts (termed *enzymes*) that promote chemical reactions. Like DNA and RNA, proteins are linear polymers. However, proteins are more complicated in that they are formed from a selection of 20 building blocks, called *amino acids*, rather than 4.

The functional properties of proteins, like those of other biomolecules, are determined by their three-dimensional structures. Proteins possess an extremely important property: a protein spontaneously folds into a welldefined and elaborate three-dimensional structure that is dictated entirely by the sequence of amino acids along its chain (Figure 1.6). *The self-folding nature of proteins constitutes the transition from the one-dimensional world of sequence information to the three-dimensional world of biological function*. This marvelous ability of proteins to self assemble into complex

structures is responsible for their dominant role in biochemistry.

How is the sequence of bases along DNA translated into a sequence of amino acids along a protein chain? We will consider the details of this process in later chapters, but the important finding is that *three bases along a DNA chain encode a single amino acid*. The specific correspondence between a set of three bases and 1 of the 20 amino acids is called the *genetic code*. Like the use of DNA as the genetic material, the genetic code is essentially universal; the same sequences of three bases encode the same amino acids in all life forms from simple microorganisms to complex, multicellular organisms such as human beings.

Knowledge of the functional and structural properties of proteins is absolutely essential to understanding the significance of the human genome sequence. For example, the sequence at the beginning of this chapter corresponds to a region of the genome that differs in people who have the genetic disorder *cystic fibrosis*. The most common mutation causing cystic fibrosis, the loss of three consecutive Ts from the gene sequence, leads to the loss of a single amino acid within a protein chain of 1480 amino acids. This seemingly slight difference—a loss of 1 amino acid of nearly 1500—creates a life-threatening condition. What is the normal function of the protein encoded by this gene? What properties of the encoded protein are compromised by this subtle defect? Can this knowledge be used to develop new treatments? These questions fall in the realm of biochemistry. Knowledge of the human genome sequence will greatly accelerate the pace at which connections are made between DNA sequences and disease as well as other human characteristics. However, these connections will be nearly meaningless without the knowledge of biochemistry necessary to interpret and exploit them.

Cystic fibrosis-

A disease that results from a decrease in fluid and salt secretion by a transport protein referred to as the cystic fibrosis transmembrane conductance regulator (CFTR). As a result of this defect, secretion from the pancreas is blocked, and heavy, dehydrated mucus accumulates in the lungs, leading to chronic lung infections.

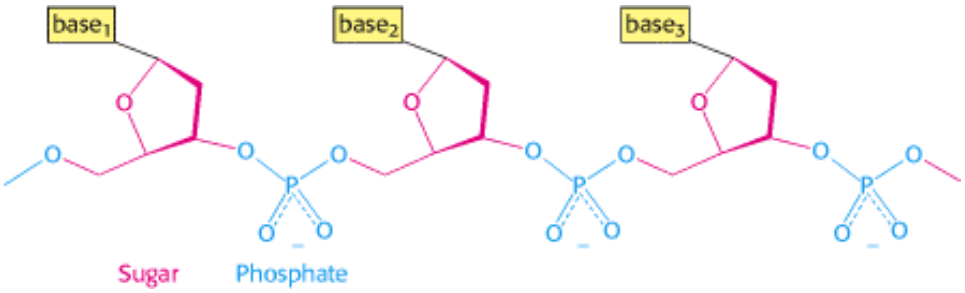


Figure 1.1. Covalent Structure of DNA. Each unit of the polymeric structure is composed of a sugar (deoxyribose), a phosphate, and a variable base that protrudes from the sugar-phosphate backbone.

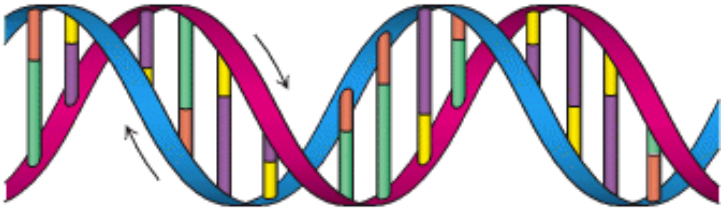


Figure 1.2. The Double Helix. The double-helical structure of DNA proposed by Watson and Crick. The sugar-phosphate backbones of the two chains are shown in red and blue and the bases are shown in green, purple, orange, and yellow.

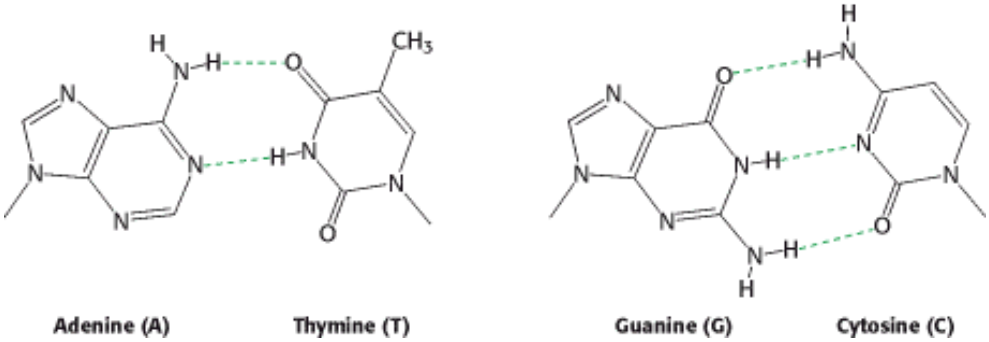


Figure 1.3. Watson-Crick Base Pairs. Adenine pairs with thymine (A-T), and guanine with cytosine (G-C). The dashed lines represent hydrogen bonds.

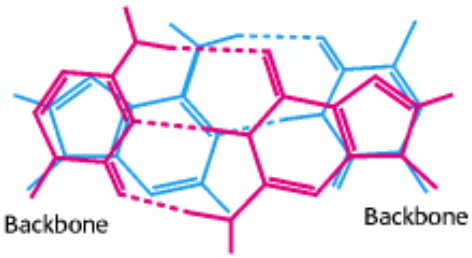


Figure 1.4. Base-Pairing in DNA. The base-pairs A-T (blue) and C-G (red) are shown overlaid. The Watson-Crick base-pairs have the same overall size and shape, allowing them to fit neatly within the double helix.

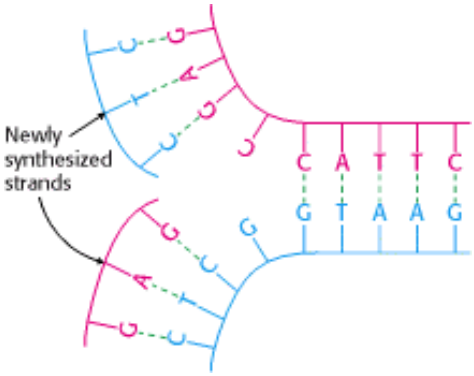


Figure 1.5. DNA Replication. If a DNA molecule is separated into two strands, each strand can act as the template for the generation of its partner strand.

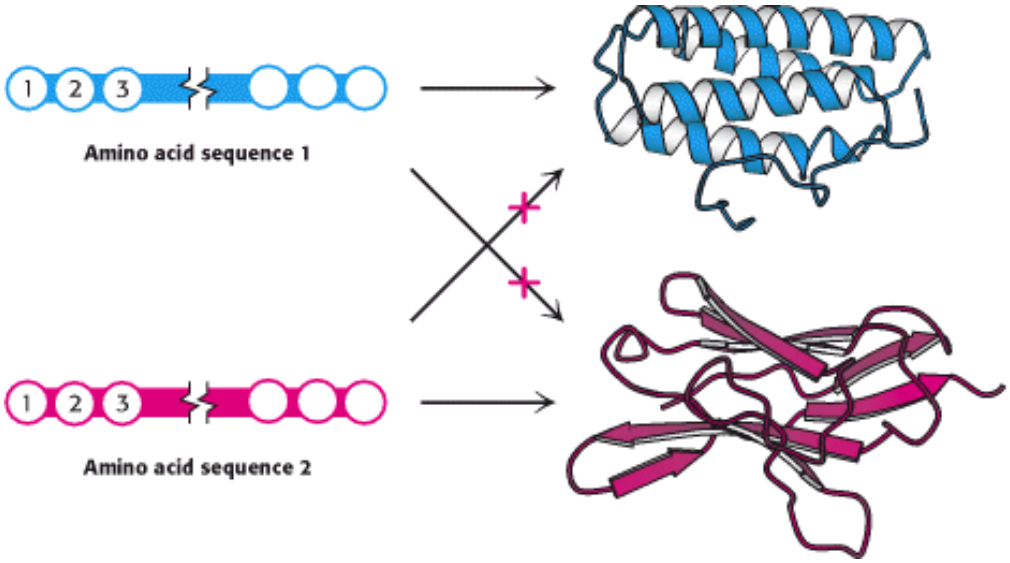


Figure 1.6. Folding of a Protein. The three-dimensional structure of a protein, a linear polymer of amino acids, is dictated by its amino acid sequence.

1.2. Biochemical Unity Underlies Biological Diversity

The stunning variety of living systems (Figure 1.7) belies a striking similarity. The common use of DNA and the genetic code by all organisms underlies one of the most powerful discoveries of the past century—namely, that *organisms are remarkably uniform at the molecular level*. All organisms are built from similar molecular components distinguishable by relatively minor variations. *This uniformity reveals that all organisms on Earth have arisen from a common ancestor*. A core of essential biochemical processes, common to all organisms, appeared early in the evolution of life. The diversity of life in the modern world has been generated by evolutionary processes acting on these core processes through millions or even billions of years. As we will see repeatedly, the generation of diversity has very often resulted from the adaptation of existing biochemical components to new roles rather than the development of fundamentally new biochemical technology. The striking uniformity of life at the molecular level affords the student of biochemistry a particularly clear view into the essence of biological processes that applies to all organisms from human beings to the simplest microorganisms.

On the basis of their biochemical characteristics, the diverse organisms of the modern world can be divided into three fundamental groups called *domains: Eukarya* (eukaryotes), *Bacteria* (formerly Eubacteria), and *Archaea* (formerly Archaebacteria). Eukarya comprise all macroscopic organisms, including human beings as well as many microscopic, unicellular organisms such as yeast. The defining characteristic of *eukaryotes* is the presence of a well-defined nucleus within each cell. Unicellular organisms such as bacteria, which lack a nucleus, are referred to as *prokaryotes*. The prokaryotes were reclassified as two separate domains in response to Carl Woese's discovery in 1977 that certain bacteria-like organisms are biochemically quite distinct from better-characterized bacterial species. These organisms, now recognized as having diverged from bacteria early in evolution, are archaea. Evolutionary paths from a common ancestor to modern organisms can be developed and analyzed on the basis of biochemical information. One such path is shown in Figure 1.8.

By examining biochemistry in the context of the tree of life, we can often understand how particular molecules or processes helped organisms adapt to specific environments or life styles. We can ask not only *what* biochemical processes take place, but also *why* particular strategies appeared in the course of evolution. In addition to being sources of historical insights, *the answers to such questions are often highly instructive with regard to the biochemistry of contemporary organisms*.

(A)

(B)

(C)

Figure 1.7. The Diversity of Living Systems. The distinct morphologies of the three organisms shown-a plant (the false hellebora, or Indian poke) and two animals (sea urchins and a common house cat)-might suggest that they have little in common. Yet biochemically they display a remarkable commonality that attests to a common ancestry. [(Left and right) John Dudak/Phototake. (Middle) Jeffrey L. Rotman/Peter Arnold.]

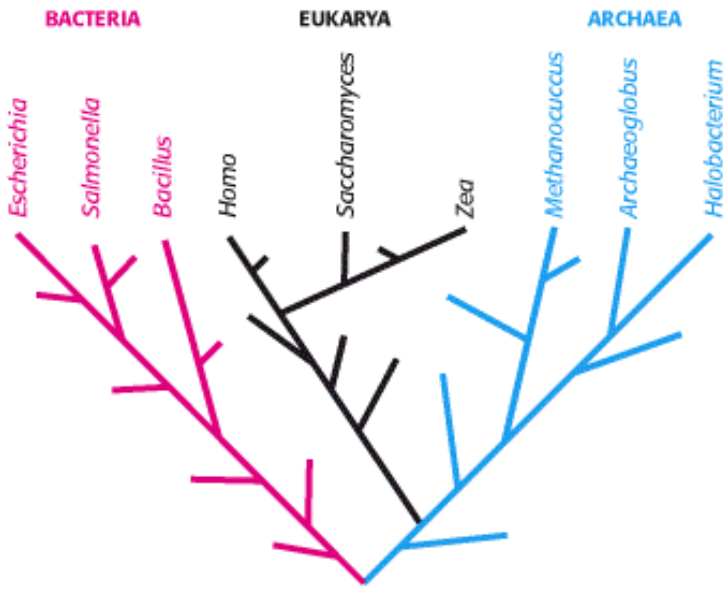


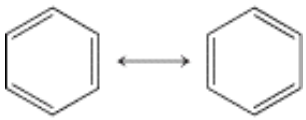
Figure 1.8. The Tree of Life. A possible evolutionary path from a common ancestral cell to the diverse species present in the modern world can be deduced from DNA sequence analysis.

1.3. Chemical Bonds in Biochemistry

The essence of biological processes—the basis of the uniformity of living systems—is in its most fundamental sense molecular interactions; in other words, the chemistry that takes place between molecules. Biochemistry is the *chemistry* that takes place within living systems. To truly understand biochemistry, we need to understand chemical bonding. We review here the types of chemical bonds that are important for biochemicals and their transformations.

The strongest bonds that are present in biochemicals are *covalent bonds*, such as the bonds that hold the atoms together within the individual bases shown in Figure 1.3. A covalent bond is formed by the sharing of a pair of electrons between adjacent atoms. A typical carbon-carbon (C-C) covalent bond has a bond length of 1.54 Å and bond energy of 85 kcal mol^{-1} (356 kJ mol^{-1}). Because this energy is relatively high, considerable energy must be expended to break covalent bonds. More than one electron pair can be shared between two atoms to form a multiple covalent bond. For example, three of the bases in Figure 1.4 include carbon-oxygen (C=O) double bonds. These bonds are even stronger than C-C single bonds, with energies near 175 kcal mol^{-1} (732 kJ mol^{-1}).

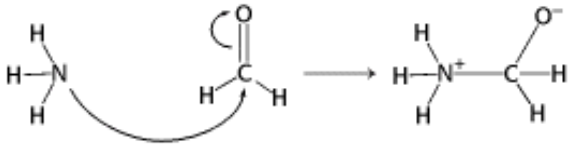
For some molecules, more than one pattern of covalent bonding can be written. For example, benzene can be written in two equivalent ways called *resonance structures*. Benzene's true structure is a composite of its two resonance structures. A molecule that can be written as several resonance structures of approximately equal energies has greater stability than does a molecule without multiple resonance structures. Thus, because of its resonance structures, benzene is unusually stable.



Benzene resonance structures

Chemical reactions entail the breaking and forming of covalent bonds. The flow of electrons in the course of a reaction can be depicted by curved arrows, a method of representation called "arrow pushing." Each arrow represents an electron

pair.

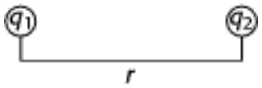


1.3.1. Reversible Interactions of Biomolecules Are Mediated by Three Kinds of Noncovalent Bonds

Readily reversible, noncovalent molecular interactions are key steps in the dance of life. Such weak, noncovalent forces play essential roles in the faithful replication of DNA, the folding of proteins into intricate three-dimensional forms, the specific recognition of substrates by enzymes, and the detection of molecular signals. Indeed, all biological structures and processes depend on the interplay of noncovalent interactions as well as covalent ones. The three fundamental noncovalent bonds are *electrostatic interactions, hydrogen bonds,* and *van der Waals interactions*. They differ in geometry, strength, and specificity. Furthermore, these bonds are greatly affected in different ways by the presence of water. Let us consider the characteristics of each:

1. *Electrostatic interactions*. An electrostatic interaction depends on the electric charges on atoms. The energy of an electrostatic interaction is given by *Coulomb's law:*

$$E = kq_1q_2/Dr$$



where E is the energy, q_1 and q_2 are the charges on the two atoms (in units of the electronic charge), r is the distance between the two atoms (in angstroms), D is the dielectric constant (which accounts for the effects of the intervening medium), and k is a proportionality constant (k = 332, to give energies in units of kilocalories per mole, or 1389, for energies in kilojoules per mole). Thus, the electrostatic interaction between two atoms bearing single opposite charges separated by 3 Å in water (which has a dielectric constant of 80) has an energy of 1.4 kcal mol^{-1} (5.9 kJ mol^{-1}).

2. *Hydrogen bonds*. Hydrogen bonds are relatively weak interactions, which nonetheless are crucial for biological macromolecules such as DNA and proteins. These interactions are also responsible for many of the properties of water that make it such a special solvent. The hydrogen atom in a hydrogen bond is partly shared between two relatively electronegative atoms such as nitrogen or oxygen. The *hydrogen-bond donor* is the group that includes both the atom to which the hydrogen is more tightly linked and the hydrogen atom itself, whereas the *hydrogen-bond acceptor* is the atom less tightly linked to the hydrogen atom (Figure 1.9). Hydrogen bonds are fundamentally electrostatic interactions. The relatively electronegative atom to which the hydrogen atom is covalently bonded pulls electron density away from the hydrogen atom so that it develops a partial positive charge (δ^+). Thus, it can interact with an atom having a partial negative charge (δ^-) through an electrostatic interaction.

Hydrogen bonds are much weaker than covalent bonds. They have energies of 1–3 kcal mol^{-1} (4–13 kJ mol^{-1}) compared

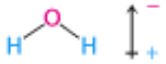
with approximately 100 kcal mol^{-1} (418 kJ mol^{-1}) for a carbon-hydrogen covalent bond. Hydrogen bonds are also somewhat longer than are covalent bonds; their bond distances (measured from the hydrogen atom) range from 1.5 to 2.6 Å; hence, distances ranging from 2.4 to 3.5 Å separate the two nonhydrogen atoms in a hydrogen bond. The strongest hydrogen bonds have a tendency to be approximately straight, such that the hydrogen-bond donor, the hydrogen atom, and the hydrogen-bond acceptor lie along a straight line.

3. *van der Waals interactions*. The basis of a van der Waals interaction is that the distribution of electronic charge around an atom changes with time. At any instant, the charge distribution is not perfectly symmetric. This transient asymmetry in the electronic charge around an atom acts through electrostatic interactions to induce a complementary asymmetry in the electron distribution around its neighboring atoms. The resulting attraction between two atoms increases as they come closer to each other, until they are separated by the van der Waals *contact distance* (Figure 1.10). At a shorter distance, very strong repulsive forces become dominant because the outer electron clouds overlap.

Energies associated with van der Waals interactions are quite small; typical interactions contribute from 0.5 to 1.0 kcal mol^{-1} (from 2 to 4 kJ mol^{-1}) per atom pair. When the surfaces of two large molecules come together, however, a large number of atoms are in van der Waals contact, and the net effect, summed over many atom pairs, can be substantial.

1.3.2. The Properties of Water Affect the Bonding Abilities of Biomolecules

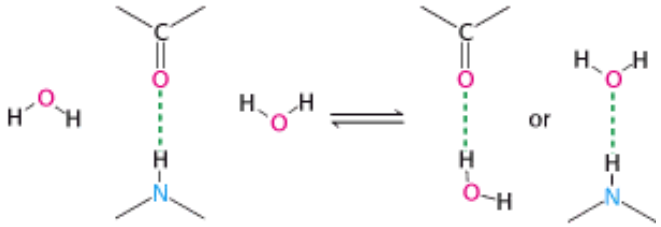
Weak interactions are the key means by which molecules interact with one another—enzymes with their substrates, hormones with their receptors, antibodies with their antigens. The strength and specificity of weak interactions are highly dependent on the medium in which they take place, and the majority of biological interactions take place in water. Two properties of water are especially important biologically:



1. *Water is a polar molecule*. The water molecule is bent, not linear, and so the distribution of charge is asymmetric. The oxygen nucleus draws electrons away from the hydrogen nuclei, which leaves the region around the hydrogen nuclei with a net positive charge. The water molecule is thus an electrically polar structure.

2. *Water is highly cohesive*. Water molecules interact strongly with one another through hydrogen bonds. These interactions are apparent in the structure of ice (Figure 1.11). Networks of hydrogen bonds hold the structure together; simi-lar interactions link molecules in liquid water and account for the cohesion of liquid water, although, in the liquid state, some of the hydrogen bonds are broken. The highly cohesive nature of water dramatically affects the interactions between molecules in aqueous solution.

What is the effect of the properties of water on the weak interactions discussed in Section 1.3.1? The polarity and hydrogen-bonding capability of water make it a highly interacting molecule. Water is an excellent solvent for polar molecules. The reason is that water greatly weakens electrostatic forces and hydrogen bonding between polar molecules by competing for their attractions. For example, consider the effect of water on hydrogen bonding between a carbonyl group and the NH group of an amide.



A hydrogen atom of water can replace the amide hydrogen atom as a hydrogen-bond donor, whereas the oxygen atom of water can replace the carbonyl oxygen atom as a hydrogen-bond acceptor. Hence, a strong hydrogen bond between a CO group and an NH group forms only if water is excluded.

The dielectric constant of water is 80, so water diminishes the strength of electrostatic attractions by a factor of 80 compared with the strength of those same interactions in a vacuum. The dielectric constant of water is unusually high because of its polarity and capacity to form oriented solvent shells around ions. These oriented solvent shells produce electric fields of their own, which oppose the fields produced by the ions. Consequently, the presence of water markedly weakens electrostatic interactions between ions.

The existence of life on Earth depends critically on the capacity of water to dissolve a remarkable array of polar molecules that serve as fuels, building blocks, catalysts, and information carriers. High concentrations of these polar molecules can coexist in water, where they are free to diffuse and interact with one another. However, the excellence of water as a solvent poses a problem, because it also weakens interactions between polar molecules. *The presence of water-free microenvironments within biological systems largely circumvents this problem*. We will see many examples of these specially constructed niches in protein molecules. Moreover, the presence of water with its polar nature permits another kind of weak interaction to take place, one that drives the folding of proteins (Section 1.3.4) and the formation of cell boundaries (Section 12.4).

The essence of these interactions, like that of all interactions in biochemistry, is energy. To understand much of biochemistry—bond formation, molecular structure, enzyme catalysis—we need to understand energy. Thermodynamics provides a valuable tool for approaching this topic. We will revisit this topic in more detail when we consider enzymes (Chapter 8) and the basic concepts of metabolism (Chapter 14).

1.3.3. Entropy and the Laws of Thermodynamics

The highly structured, organized nature of living organisms is apparent and astonishing. This organization extends from the organismal through the cellular to the molecular level. Indeed, biological processes can seem magical in that the well-ordered structures and patterns emerge from the chaotic and disordered world of inanimate objects. However, the organization visible in a cell or a molecule arises from biological events that are subject to the same physical laws that govern all processes—in particular, the *laws of thermodynamics*.

How can we understand the creation of order out of chaos? We begin by noting that the laws of thermodynamics make a distinction between a system and its surroundings. A *system* is defined as the matter within a defined region of space. The matter in the rest of the universe is called the *surroundings. The First Law of Thermodynamics states that the total energy of a system and its surroundings is constant*. In other words, the energy content of the universe is constant; energy can be neither created nor destroyed. Energy can take different forms, however. Heat, for example, is one form of energy. Heat is a manifestation of the *kinetic energy* associated with the random motion of molecules. Alternatively, energy can be present as *potential energy*, referring to the ability of energy to be released on the occurrence of some process. Consider, for example, a ball held at the top of a tower. The ball has considerable potential energy because, when it is released, the ball will develop kinetic energy associated with its motion as it falls. Within chemical systems, potential energy is related to the likelihood that atoms can react with one another. For instance, a mixture of gasoline and oxygen has much potential energy because these molecules may react to form carbon dioxide and release energy as heat. The First Law requires that any energy released in the formation of chemical bonds be used to break other bonds, be released as heat, or be stored in some other form.

Another important thermodynamic concept is that of *entropy*. Entropy is a measure of the level of randomness or disorder in a system. *The Second Law of Thermodynamics states that the total entropy of a system and its surroundings always increases for a spontaneous process*. At first glance, this law appears to contradict much common experience, particularly about biological systems. Many biological processes, such as the generation of a well-defined structure such as a leaf from carbon dioxide gas and other nutrients, clearly increase the level of order and hence decrease entropy. Entropy may be decreased locally in the formation of such ordered structures only if the entropy of other parts of the universe is increased by an equal or greater amount.

An example may help clarify the application of the laws of thermodynamics to a chemical system. Consider a container with 2 moles of hydrogen gas on one side of a divider and 1 mole of oxygen gas on the other (Figure 1.12). If the divider is removed, the gases will intermingle spontaneously to form a uniform mixture. The process of mixing increases entropy as an ordered arrangement is replaced by a randomly distributed mixture.

Other processes within this system can decrease the entropy locally while increasing the entropy of the universe. A spark applied to the mixture initiates a chemical reaction in which hydrogen and oxygen combine to form water:

$$2\,H_2 + O_2 \longrightarrow 2\,H_2O$$

If the temperature of the system is held constant, the entropy of the system decreases because 3 moles of two differing reactants have been combined to form 2 moles of a single product. The gas now consists of a uniform set of indistinguishable molecules. However, the reaction releases a significant amount of heat into the surroundings, and this heat will increase the entropy of the surrounding molecules by increasing their random movement. The entropy increase in the surroundings is enough to allow water to form spontaneously from hydrogen and oxygen (Figure 1.13).

The change in the entropy of the surroundings will be proportional to the amount of heat transferred from the system and inversely proportional to the temperature of the surroundings, because an input of heat leads to a greater increase in entropy at lower temperatures than at higher temperatures. In biological systems, *T* [in kelvin (K), absolute temperature] is assumed to be constant. If we define the heat content of a system as *enthalpy (H)*, then we can express the relation linking the entropy *(S)* of the surroundings to the transferred heat and temperature as a simple equation:

$$\Delta S_{\text{surroundings}} = -\Delta H_{\text{system}}/T \qquad (1)$$

The total entropy change is given by the expression

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{surroundings}} \qquad (2)$$

Substituting equation 1 into equation 2 yields

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} - \Delta H_{\text{system}}/T \qquad (3)$$

Multiplying by -*T* gives

$$-T\Delta S_{\text{total}} = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \qquad (4)$$

The function $-T\,\Delta\,S$ has units of energy and is referred to as *free energy* or *Gibbs free energy*, after Josiah Willard Gibbs, who developed this function in 1878:

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \qquad (5)$$

The free-energy change, ΔG, will be used throughout this book to describe the energetics of biochemical reactions.

Recall that the Second Law of Thermodynamics states that, for a reaction to be spontaneous, the entropy of the universe must increase. Examination of equation 3 shows that the total entropy will increase if and only if

$$\Delta S_{\text{system}} > \Delta H_{\text{system}}/T \qquad (6)$$

Rearranging gives $T\Delta S_{\text{system}} > \Delta H$, or entropy will increase if and only if

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} < 0 \qquad (7)$$

In other words, *the free-energy change must be negative for a reaction to be spontaneous*. A negative free-energy change occurs with an increase in the overall entropy of the universe. Thus, we need to consider only one term, the free energy of the system, to decide whether a reaction can occur spontaneously; any effects of the changes within the system on the rest of the universe are automatically taken into account.

1.3.4. Protein Folding Can Be Understood in Terms of Free-Energy Changes

The problem of protein folding illustrates the utility of the concept of free energy. Consider a system consisting of a solution of unfolded protein molecules in aqueous solution (Figure 1.14). Each unfolded protein molecule can adopt a unique conformation, so the system is quite disordered and the entropy of the collection of molecules is relatively high. Yet, protein folding proceeds spontaneously under appropriate conditions. Thus, entropy must be increasing elsewhere in the system or in the surroundings. How can we reconcile the apparent contradiction that proteins spontaneously assume an ordered structure, and yet entropy increases? The entropy decrease in the system on folding is not as large as it appears to be, because of the properties of water. Molecules in aqueous solution interact with water molecules through the formation of hydrogen and ionic interactions. However, some molecules (termed *nonpolar molecules*) cannot participate in hydrogen or ionic interactions. The interactions of nonpolar molecules with water are not as favorable as are interactions between the water molecules themselves. The water molecules in contact with these nonpolar surfaces form "cages" around the nonpolar molecule, becoming more well ordered (and, hence, lower in entropy) than water molecules free in solution. As two such nonpolar molecules come together, some of the water molecules are released, and so they can interact freely with bulk water (Figure 1.15). Hence, nonpolar molecules have a tendency to aggregate in water because the entropy of the water is increased through the release of water molecules. This phenomenon, termed the *hydrophobic effect*, helps promote many biochemical processes.

How does the hydrophobic effect favor protein folding? Some of the amino acids that make up proteins have nonpolar groups. These nonpolar amino acids have a strong tendency to associate with one another inside the interior of the folded protein. The increased entropy of water resulting from the interaction of these hydrophobic amino acids helps to compensate for the entropy losses inherent in the folding process.

Hydrophobic interactions are not the only means of stabilizing protein structure. Many weak bonds, including hydrogen bonds and van der Waals interactions, are formed in the protein-folding process, and heat is released into the surroundings as a consequence. Although these interactions replace interactions with water that take place in the unfolded protein, the net result is the release of heat to the surroundings and thus a negative (favorable) change in enthalpy for the system.

The folding process can occur when the combination of the entropy associated with the hydrophobic effect and the enthalpy change associated with hydrogen bonds and van der Waals interactions makes the overall free energy negative.

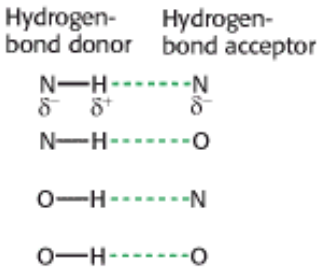


Figure 1.9. Hydrogen Bonds that Include Nitrogen and Oxygen Atoms. The positions of the partial charges (δ^+ and δ^-) are shown.

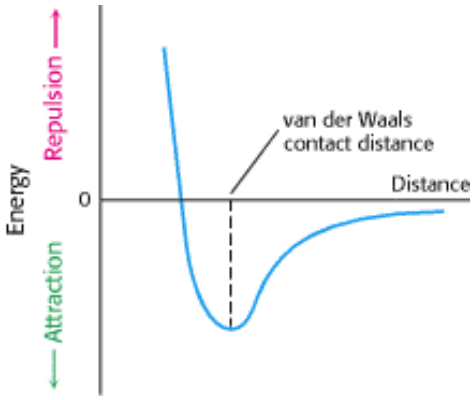


Figure 1.10. Energy of a van der Waals Interaction as Two Atoms Approach One Another. The energy is most favorable at the van der Waals contact distance. The energy rises rapidly owing to electron- electron repulsion as the atoms move closer together than this distance.

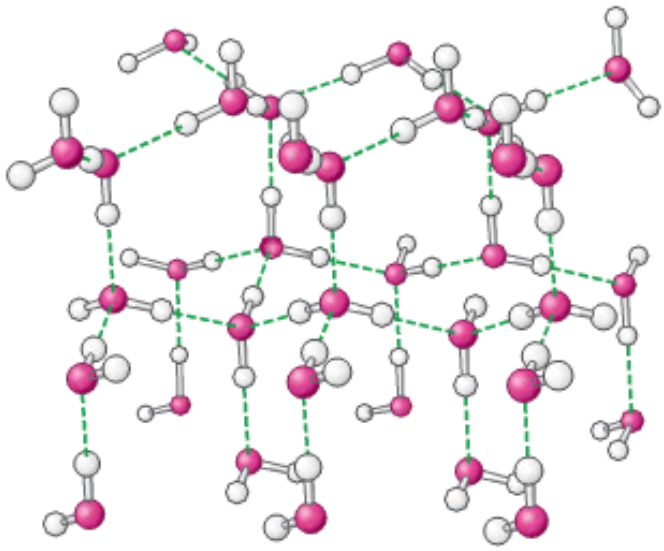


Figure 1.11. Structure of Ice. Hydrogen bonds (shown as dashed lines) are formed between water molecules.

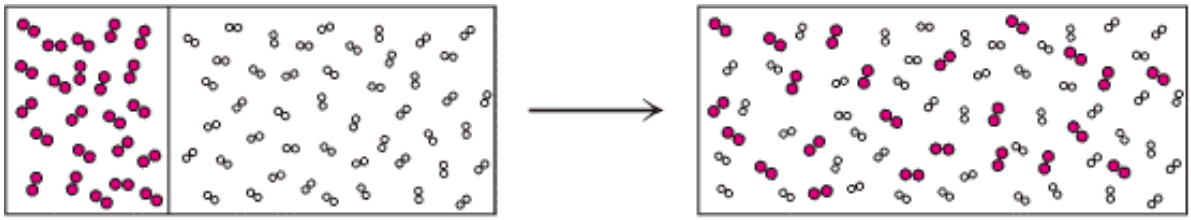


Figure 1.12. From Order to Disorder. The spontaneous mixing of gases is driven by an increase in entropy.

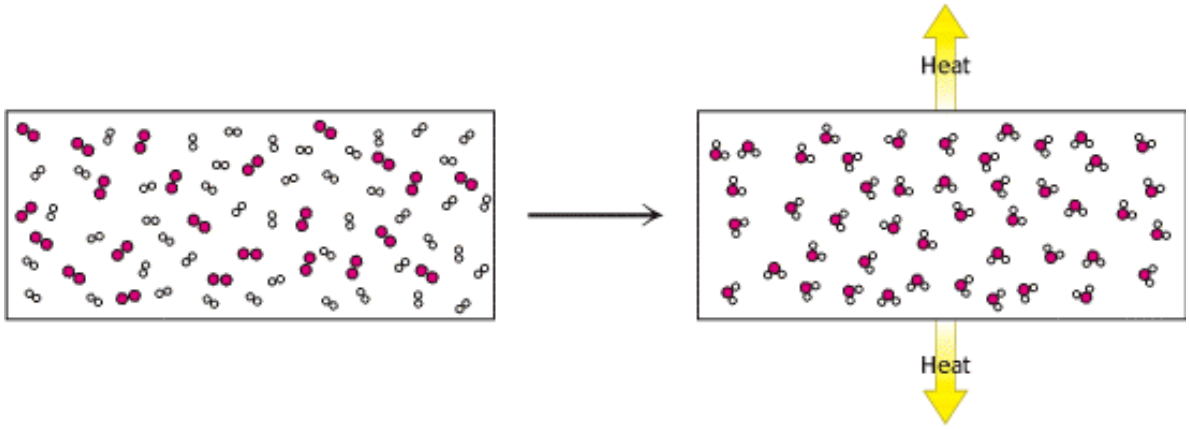


Figure 1.13. Entropy Changes. When hydrogen and oxygen combine to form water, the entropy of the system is reduced, but the entropy of the universe is increased owing to the release of heat to the surroundings.

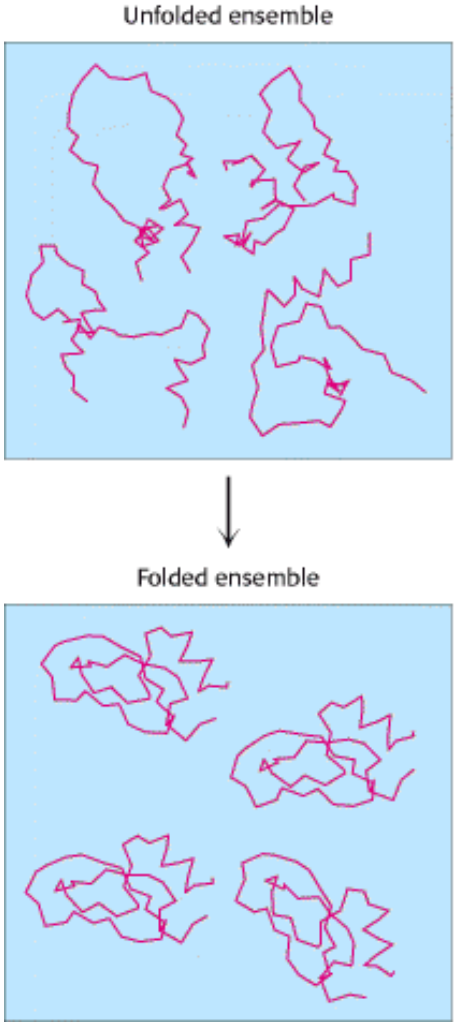


Figure 1.14. Protein Folding. Protein folding entails the transition from a disordered mixture of unfolded molecules to a relatively uniform solution of folded protein molecules.

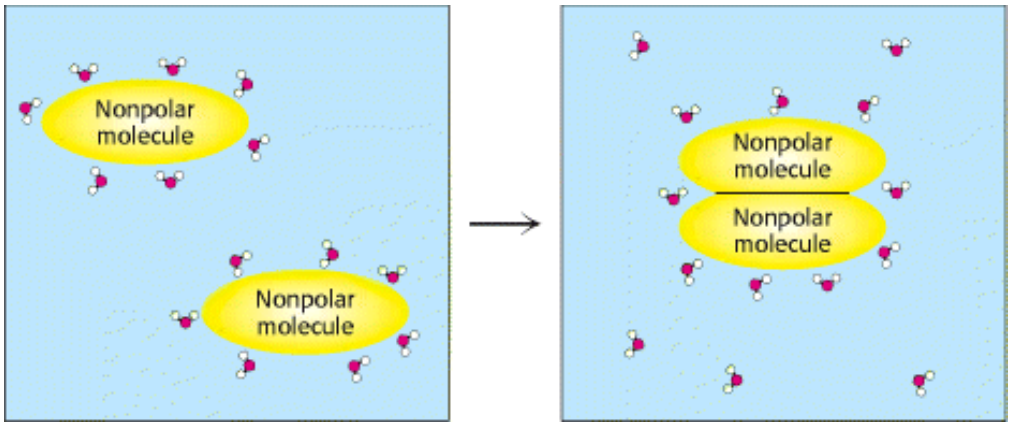


Figure 1.15. The Hydrophobic Effect. The aggregation of nonpolar groups in water leads to an increase in entropy owing to the release of water molecules into bulk water.

1.4. Biochemistry and Human Biology

Our understanding of biochemistry has had and will continue to have extensive effects on many aspects of human endeavor. *First, biochemistry is an intrinsically beautiful and fascinating body of knowledge*. We now know the essence and many of the details of the most fundamental processes in biochemistry, such as how a single molecule of DNA replicates to generate two identical copies of itself and how the sequence of bases in a DNA molecule determines the sequence of amino acids in an encoded protein. Our ability to describe these processes in detailed, mechanistic terms places a firm chemical foundation under other biological sciences. Moreover, the realization that we can understand essential life processes, such as the transmission of hereditary information, as chemical structures and their reactions has significant philosophical implications. What does it mean, biochemically, to be human? What are the biochemical differences between a human being, a chimpanzee, a mouse, and a fruit fly? Are we more similar than we are different?

Second, biochemistry is greatly influencing medicine and other fields. The molecular lesions causing sickle-cell anemia, cystic fibrosis, hemophilia, and many other genetic diseases have been elucidated at the biochemical level. Some of the molecular events that contribute to cancer development have been identified. An understanding of the underlying defects opens the door to the discovery of effective therapies. Biochemistry makes possible the rational design of new drugs, including specific inhibitors of enzymes required for the replication of viruses such as human immunodeficiency virus (HIV). Genetically engineered bacteria or other organisms can be used as "factories" to produce valuable proteins such as insulin and stimulators of blood-cell development. Biochemistry is also contributing richly to clinical diagnostics. For example, elevated levels of telltale enzymes in the blood reveal whether a patient has recently had a myocardial infarction (heart attack). DNA probes are coming into play in the precise diagnosis of inherited disorders, infectious diseases, and cancers. Agriculture, too, is benefiting from advances in biochemistry with the development of more effective, environmentally safer herbicides and pesticides and the creation of genetically engineered plants that are, for example, more resistant to insects. All of these endeavors are being accelerated by the advances in genomic sequencing.

Third, advances in biochemistry are enabling researchers to tackle some of the most exciting questions in biology and medicine. How does a fertilized egg give rise to cells as different as those in muscle, brain, and liver? How do the senses work? What are the molecular bases for mental disorders such as Alzheimer disease and schizophrenia? How does the immune system distinguish between self and nonself? What are the molecular mechanisms of short-term and long-term memory? The answers to such questions, which once seemed remote, have been partly uncovered and are likely to be more thoroughly revealed in the near future.

Because all living organisms on Earth are linked by a common origin, evolution provides a powerful organizing theme for biochemistry. This book is organized to emphasize the unifying principles revealed by evolutionary considerations. We begin in the next chapter with a brief tour along a plausible evolutionary path from the formation of some of the chemicals that we now associate with living organisms through the evolution of the processes essential for the development of complex, multicellular organisms. The remainder of Part I of the book more fully introduces the most important classes of biochemicals as well as catalysis and regulation. Part II, Transducing and Storing Energy, describes how energy from chemicals or from sunlight is converted into usable forms and how this conversion is regulated. As we will see, a small set of molecules such as adenosine triphosphate (ATP) act as energy currencies that allow energy, however captured, to be utilized in a variety of biochemical processes. This part of the text examines the important pathways for the conversion of environmental energy into molecules such as ATP and uncovers many unifying principles. Part III, Synthesizing the Molecules of Life, illustrates the use of the molecules discussed in Part II to synthesize key molecular building blocks, such as the bases of DNA and amino acids, and then shows how these precursors are assembled into DNA, RNA, and proteins. In Parts II and III, we will highlight the relation between the reactions within each pathway and between those in different pathways so as to suggest how these individual reactions may have combined early in evolutionary history to produce the necessary molecules. From the student's perspective, the existence of features common to several pathways enables material mastered in one context to be readily applied to new contexts. Part IV, Responding to Environmental Changes, explores some of the mechanisms that cells and multicellular organisms have evolved to detect and respond to changes in the environment. The topics range from general mechanisms, common to all organisms, for regulating the expression of genes to the sensory systems used by human

beings and other complex organisms. In many cases, we can now see how these elaborate systems evolved from pathways that existed earlier in evolutionary history. Many of the sections in Part IV link biochemistry with other fields such as cell biology, immunology, and neuroscience. We are now ready to begin our journey into biochemistry with events that took place more than 3 billion years ago.

Appendix: Depicting Molecular Structures

The authors of a biochemistry text face the problem of trying to present three-dimensional molecules in the two dimensions available on the printed page. The interplay between the three-dimensional structures of biomolecules and their biological functions will be discussed extensively throughout this book. Toward this end, we will frequently use representations that, although of necessity are rendered in two dimensions, emphasize the three-dimensional structures of molecules.

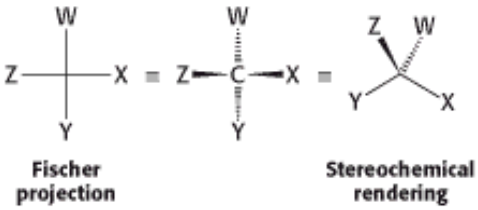
Stereochemical Renderings

Most of the chemical formulas in this text are drawn to depict the geometric arrangement of atoms, crucial to chemical bonding and reactivity, as accurately as possible. For example, the carbon atom of methane is sp^3 hybridized and tetrahedral, with H-C-H angles of 109.5 degrees while the carbon atom in formaldehyde is sp^2 hybridized with bond angles of 120 degrees.

To illustrate the correct *stereochemistry* about carbon atoms, wedges will be used to depict the direction of a bond into or out of the plane of the page. A solid wedge with the broad end away from the carbon denotes a bond coming toward the viewer out of the plane. A dashed wedge, with the broad end of the bond at the carbon represents a bond going away from the viewer into the plane of the page. The remaining two bonds are depicted as straight lines.

Fischer Projections

Although more representative of the actual structure of a compound, stereochemical structures are often difficult to draw quickly. An alternative method of depicting structures with tetrahedral carbon centers relies on the use of *Fischer projections*.



In a Fischer projection, the bonds to the central carbon are represented by horizontal and vertical lines from the substituent atoms to the carbon atom, which is assumed to be at the center of the cross. By convention, the horizontal bonds are assumed to project out of the page toward the viewer, whereas the vertical bonds are assumed to project into the page away from the viewer. The Glossary of Compounds found at the back of the book is a structural glossary of the

key molecules in biochemistry, presented both as stereochemically accurate structures and as Fisher projections.

For depicting molecular architecture in more detail, five types of models will be used: space filling, ball and stick, skeletal, ribbon, and surface representations (Figure 1.16). The first three types show structures at the atomic level.

1. *Space-filling models*. The space-filling models are the most realistic. The size and position of an atom in a space-filling model are determined by its bonding properties and van der Waals radius, or contact distance (Section 1.3.1). A van der Waals radius describes how closely two atoms can approach each other when they are not linked by a covalent bond. The colors of the model are set by convention.

Carbon, black	Hydrogen, white	Nitrogen, blue
Oxygen, red	Sulfur, yellow	Phosphorus, purple

Space-filling models of several simple molecules are shown in Figure 1.17.

2. *Ball-and-stick models*. Ball-and-stick models are not as realistic as space-filling models, because the atoms are depicted as spheres of radii smaller than their van der Waals radii. However, the bonding arrangement is easier to see because the bonds are explicitly represented as sticks. In an illustration, the taper of a stick, representing parallax, tells which of a pair of bonded atoms is closer to the reader. A ball-and-stick model reveals a complex structure more clearly than a space-filling model does.

3. *Skeletal models*. An even simpler image is achieved with a skeletal model, which shows only the molecular framework. In skeletal models, atoms are not shown explicitly. Rather, their positions are implied by the junctions and ends of bonds. Skeletal models are frequently used to depict larger, more complex structures.

As biochemistry has advanced, more attention has been focused on the structures of biological macromolecules and their complexes. These structures comprise thousands or even tens of thousands of atoms. Although these structures can be depicted at the atomic level, it is difficult to discern the relevant structural features because of the large number of atoms. Thus, more schematic representations—ribbon diagrams and surface representations—have been developed for the depiction of macromolecular structures in which atoms are not shown explicitly (Figure 1.18).

4. *Ribbon diagrams*. These diagrams are highly schematic and most commonly used to accent a few dramatic aspects of protein structure, such as the α helix (a coiled ribbon), the β strand (a broad arrow), and loops (simple lines), so as to provide simple and clear views of the folding patterns of proteins.

5. *Surface representations*. Often, the interactions between macromolecules take place exclusively at their surfaces. Surface representations have been developed to better visualize macromolecular surfaces. These representations display the overall shapes of macromolecules and can be shaded or colored to indicate particular features such as surface topography or the distribution of electric charges.

Key Terms

deoxyribonucleic acid (DNA)

double helix

ribonucleic acid (RNA)

protein

amino acid

genetic code

Eukarya

Bacteria

Archaea

eukaryote

prokaryote

covalent bond

resonance structure

electrostatic interaction

hydrogen bond

van der Waals interaction

entropy

enthalpy

free energy

hydrophobic effect

sterochemistry

Fischer projection

space-filling model

ball-and stick-model

skeletal model

ribbon diagram

surface presentation

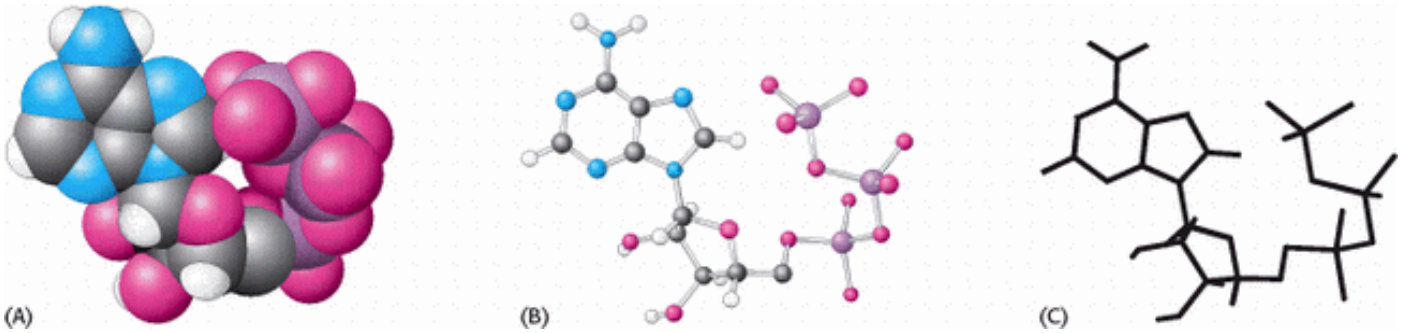


Figure 1.16. Molecular Representations. Comparison of (A) space-filling, (B) ball-and-stick, and (C) skeletal models of ATP.

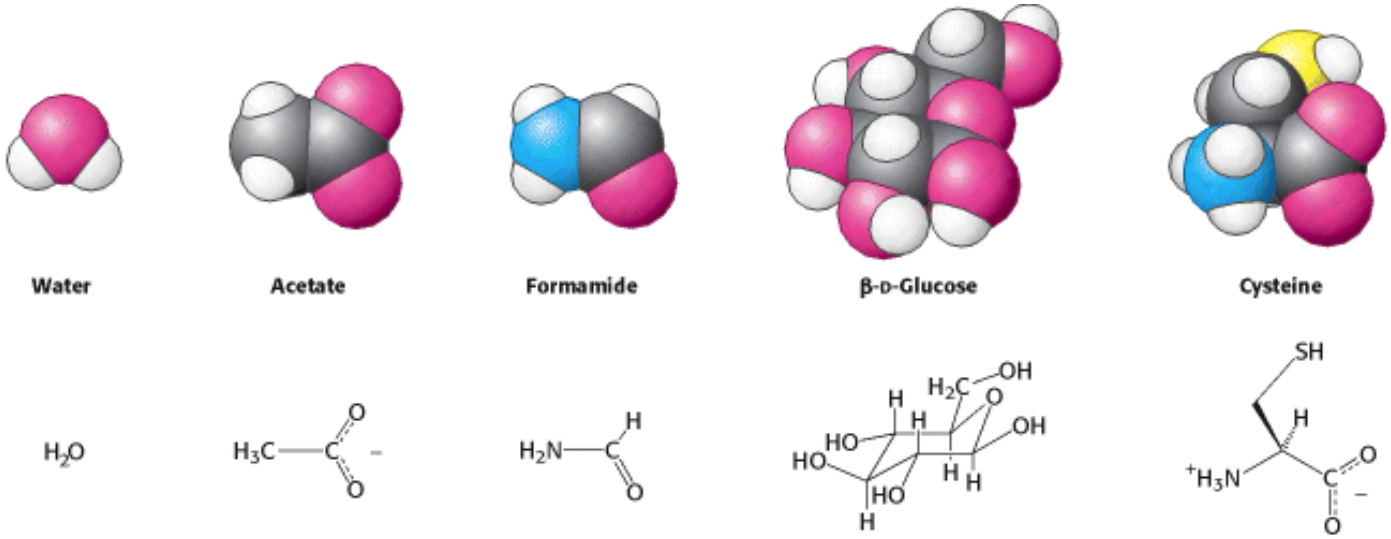


Figure 1.17. Space-Filling Models. Structural formulas and space-filling representations of selected molecules are shown.

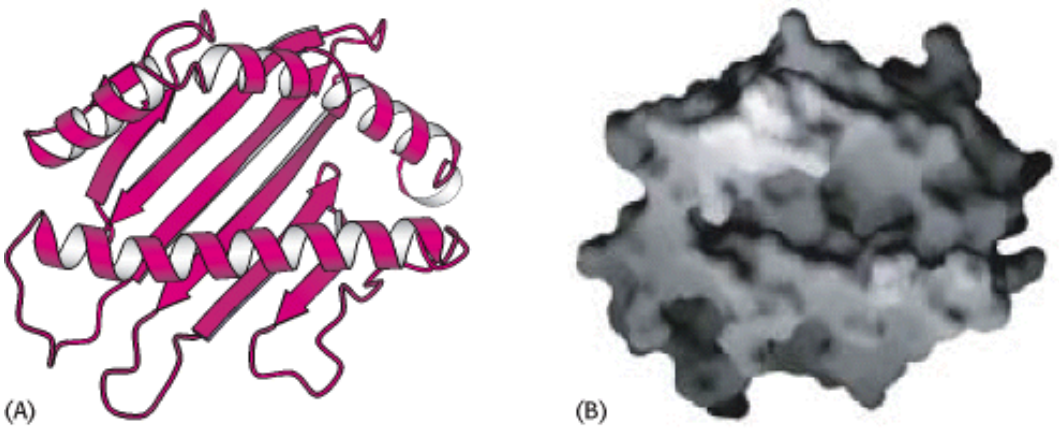


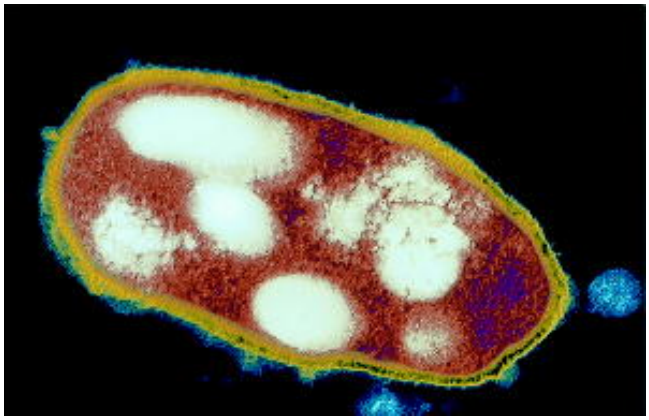
Figure 1.18. Alternative Representations of Protein Structure. A ribbon diagram (A) and a surface representation (B) of a key protein from the immune system emphasize different aspects of structure.

2. Biochemical Evolution

Earth is approximately 4.5 billion years old. Remarkably, there is convincing fossil evidence that organisms morphologically (and very probably biochemically) resembling certain modern bacteria were in existence 3.5 billion years ago. With the use of the results of directed studies and accidental discoveries, it is now possible to construct a hypothetical yet plausible evolutionary path from the prebiotic world to the present. A number of uncertainties remain, particularly with regard to the earliest events. Nonetheless, a consideration of the steps along this path and the biochemical problems that had to be solved provides a useful perspective from which to regard the processes found in modern organisms. *These evolutionary connections make many aspects of biochemistry easier to understand.*

We can think of the path leading to modern living species as consisting of stages, although it is important to keep in mind that these stages were almost certainly not as distinct as presented here. The first stage was the initial generation of some of the key molecules of life—nucleic acids, proteins, carbohydrates, and lipids—by nonbiological processes. The second stage was fundamental—the transition from prebiotic chemistry to replicating systems. With the passage of time, these systems became increasingly sophisticated, enabling the formation of living cells. In the third stage, mechanisms evolved for interconverting energy from chemical sources and sunlight into forms that can be utilized to drive biochemical reactions. Intertwined with these energy-conversion processes are pathways for synthesizing the components of nucleic acids, proteins, and other key substances from simpler molecules. With the development of energy-conversion processes and biosynthetic pathways, a wide variety of unicellular organisms evolved. The fourth stage was the evolution of mechanisms that allowed cells to adjust their biochemistry to different, and often changing, environments. Organisms with these capabilities could form colonies comprising groups of interacting cells, and some eventually evolved into complex multicellular organisms.

This chapter introduces key challenges posed in the evolution of life, whose solutions are elaborated in later chapters. Exploring a possible evolutionary origin for these fundamental processes makes their use, in contrast with that of potential alternatives, more understandable.



Natural selection, one of the key forces powering evolution, opens an array of improbable ecological niches to species that can adapt biochemically. (Left) Salt pools, where the salt concentration can be greater than 1.5 M, would seem to be highly inhospitable environments for life. Yet certain halophilic archaea, such as *Haloferax mediterranei* (right), possess biochemical adaptations that enable them to thrive under these harsh conditions. [(Left) Kaj R. Svensson/ Science Photo Library/Photo Researchers; (right) Wanner/Eye of Science/Photo Researchers.]

2.1. Key Organic Molecules Are Used by Living Systems

Approximately 1 billion years after Earth's formation, life appeared, as already mentioned. Before life could exist, though, another major process needed to have taken place—the synthesis of the organic molecules required for living systems from simpler molecules found in the environment. The components of nucleic acids and proteins are relatively complex organic molecules, and one might expect that only sophisticated synthetic routes could produce them. However, this requirement appears not to have been the case. How did the building blocks of life come to be?

2.1.1. Many Components of Biochemical Macromolecules Can Be Produced in Simple, Prebiotic Reactions

Among several competing theories about the conditions of the *prebiotic world*, none is completely satisfactory or problem-free. One theory holds that Earth's early atmosphere was highly reduced, rich in methane (CH_4), ammonia (NH_3), water (H_2O), and hydrogen (H_2), and that this atmosphere was subjected to large amounts of solar radiation and lightning. For the sake of argument, we will assume that these conditions were indeed those of prebiotic Earth. Can complex organic molecules be synthesized under these conditions? In the 1950s, Stanley Miller and Harold Urey set out to answer this question. An electric discharge, simulating lightning, was passed through a mixture of methane, ammonia, water, and hydrogen (Figure 2.1). Remarkably, these experiments yielded a highly nonrandom mixture of organic compounds, including amino acids and other substances fundamental to biochemistry. The procedure produces the amino acids glycine and alanine in approximately 2% yield, depending on the amount of carbon supplied as methane. More complex amino acids such as glutamic acid and leucine are produced in smaller amounts (Figure 2.2). Hydrogen cyanide (HCN), another likely component of the early atmosphere, will condense on exposure to heat or light to produce adenine, one of the four nucleic acid bases (Figure 2.3). Other simple molecules combine to form the remaining bases. A wide array of sugars, including ribose, can be formed from formaldehyde under prebiotic conditions.

2.1.2. Uncertainties Obscure the Origins of Some Key Biomolecules

The preceding observations suggest that many of the building blocks found in biology are unusually easy to synthesize and that significant amounts could have accumulated through the action of nonbiological processes. However, it is important to keep in mind that there are many uncertainties. For instance, ribose is just one of many sugars formed under prebiotic conditions. In addition, ribose is rather unstable under possible prebiotic conditions. Futhermore, ribose occurs in two mirror-image forms, only one of which occurs in modern RNA. To circumvent those problems, the first nucleic acid-like molecules have been suggested to have been bases attached to a different backbone and only later in

evolutionary time was ribose incorporated to form nucleic acids as we know them today. Despite these uncertainties, an assortment of prebiotic molecules did arise in some fashion, and from this assortment *those with properties favorable for the processes that we now associate with life began to interact and to form more complicated compounds*. The processes through which modern organisms synthesize molecular building blocks will be discussed in Chapters 24, 25, and 26.

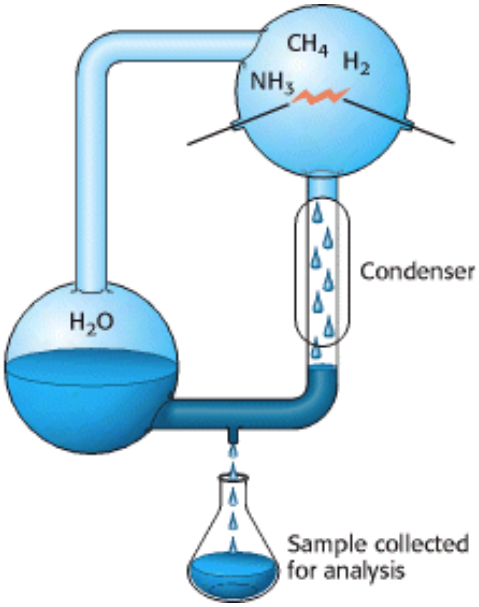


Figure 2.1. The Urey-Miller Experiment. An electric discharge (simulating lightning) passed through an atmosphere of CH_4, NH_3, H_2O, and H_2 leads to the generation of key organic compounds such as amino acids.

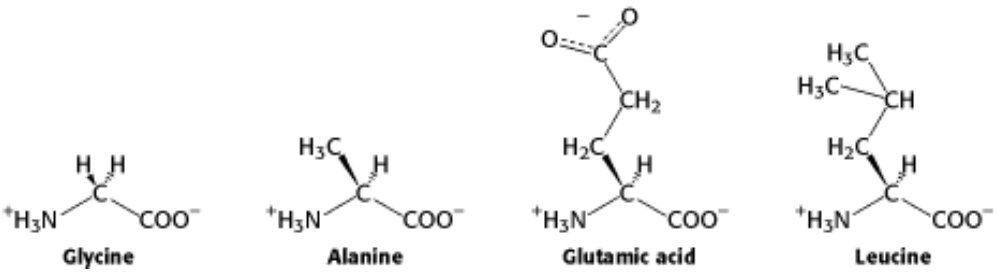


Figure 2.2. Products of Prebiotic Synthesis. Amino acids produced in the Urey-Miller experiment.

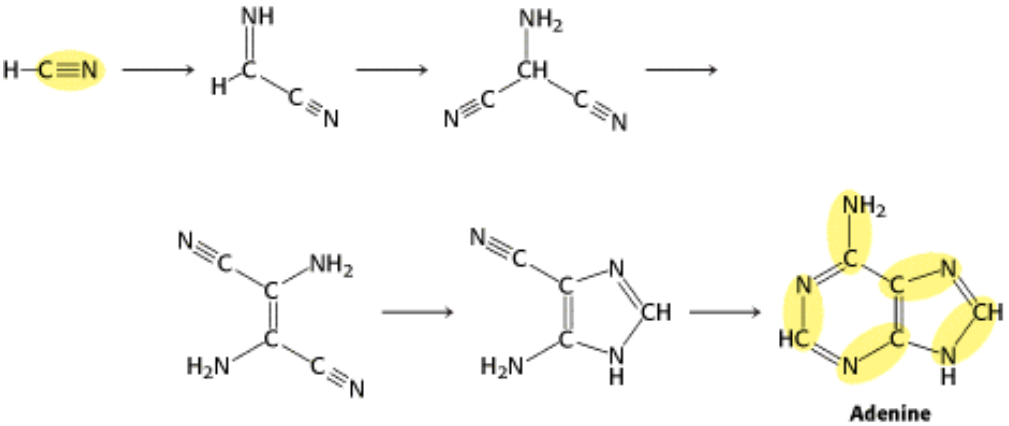


Figure 2.3. Prebiotic Synthesis of a Nucleic Acid Component. Adenine can be generated by the condensation of HCN.

2.2. Evolution Requires Reproduction, Variation, and Selective Pressure

Once the necessary building blocks were available, how did a living system arise and evolve? Before the appearance of life, simple molecular systems must have existed that subsequently evolved into the complex chemical systems that are characteristic of organisms. To address how this evolution occurred, we need to consider the *process* of evolution. There are several basic principles common to evolving systems, whether they are simple collections of molecules or competing populations of organisms. First, the most fundamental property of evolving systems is their ability to *replicate* or *reproduce*. Without this ability of *reproduction*, each "species" of molecule that might appear is doomed to extinction as soon as all its individual molecules degrade. For example, individual molecules of biological polymers such as ribonucleic acid are degraded by hydrolysis reactions and other processes. However, *molecules that can replicate will continue to be represented in the population even if the lifetime of each individual molecule remains short*.

A second principle fundamental to evolution is *variation*. The replicating systems must undergo changes. After all, if a system always replicates perfectly, the replicated molecule will always be the same as the parent molecule. Evolution cannot occur. The nature of these variations in living systems are considered in Section 2.2.5.

A third basic principle of evolution is *competition*. Replicating molecules compete with one another for available resources such as chemical precursors, and the competition allows the process of *evolution by natural selection* to occur. Variation will produce differing populations of molecules. Some variant offspring may, by chance, be better suited for survival and replication under the prevailing conditions than are their parent molecules. The prevailing conditions exert a *selective pressure* that gives an advantage to one of the variants. Those molecules that are best able to survive and to replicate themselves will increase in relative concentration. Thus, new molecules arise that are better able to replicate under the conditions of their environment. The same principles hold true for modern organisms. Organisms reproduce, show variation among individual organisms, and compete for resources; those variants with a selective advantage will reproduce more successfully. The changes leading to variation still take place at the molecular level, but the selective advantage is manifest at the organismal level.

2.2.1. The Principles of Evolution Can Be Demonstrated in Vitro

Is there any evidence that evolution can take place at the molecular level? In 1967, Sol Spiegelman showed that replicating molecules could evolve new forms in an experiment that allowed him to observe molecular evolution in the test tube. He used as his evolving molecules RNA molecules derived from a bacterial virus called bacteriophage Q β . The genome of bacteriophage Q β , a single RNA strand of approximately 3300 bases, depends for its replication on the activity of a protein complex termed Q β replicase. Spiegelman mixed the replicase with a starting population of Q β RNA molecules. Under conditions in which there are ample amounts of precursors, no time constraints, and no other selective pressures, the composition of the population does not change from that of the parent molecules on replication. When selective pressures are applied, however, the composition of the population of molecules can change dramatically. For example, decreasing the time available for replication from 20 minutes to 5 minutes yielded, incrementally over 75 generations, a population of molecules dominated by a single species comprising only 550 bases. This species is replicated 15 times as rapidly as the parental Q β RNA (Figure 2.4). Spiegelman applied other selective pressures by, for example, limiting the concentrations of precursors or adding compounds that inhibit the replication process. In each case, new species appeared that replicated more effectively under the conditions imposed.

The process of evolution demonstrated in these studies depended on the existence of machinery for the replication of RNA fragments in the form of the Q β replicase. As noted in Chapter 1, one of the most elegant characteristics of nucleic acids is that the mechanism for their replication follows naturally from their molecular structure. This observation suggests that nucleic acids, perhaps RNA, could have become *self-replicating*. Indeed, the results of studies have revealed that single-stranded nucleic acids can serve as templates for the synthesis of their complementary strands and that this synthesis can occur spontaneously—that is, without biologically derived replication machinery. However, investigators have not yet found conditions in which an RNA molecule is fully capable of independent selfreplication

from simple starting materials.

2.2.2. RNA Molecules Can Act As Catalysts

The development of capabilities beyond simple replication required the generation of specific catalysts. A *catalyst* is a molecule that accelerates a particular chemical reaction without itself being chemically altered in the process. The properties of catalysts will be discussed in detail in Chapters 8 and 9. Some catalysts are highly specific; they promote certain reactions without substantially affecting closely related processes. Such catalysts allow the reactions of specific pathways to take place in preference to those of potential alternative pathways. Until the 1980s, all biological catalysts, termed *enzymes*, were believed to be proteins. Then, Tom Cech and Sidney Altman independently discovered that certain RNA molecules can be effective catalysts. These RNA catalysts have come to be known as *ribozymes*. The discovery of ribozymes suggested the possibility that catalytic RNA molecules could have played fundamental roles early in the evolution of life.

The catalytic ability of RNA molecules is related to their ability to adopt specific yet complex structures. This principle is illustrated by a "hammerhead" ribozyme, an RNA structure first identified in plant viruses (Figure 2.5). This RNA molecule promotes the cleavage of specific RNA molecules at specific sites; this cleavage is necessary for certain aspects of the viral life cycle. The ribozyme, which requires Mg^{2+} ion or other ions for the cleavage step to take place, forms a complex with its substrate RNA molecule that can adopt a reactive conformation.

This icon, appearing throughout the book, indicates an opportunity to explore further resources available on the Biochemistry Web site; www.whfreeman.com/biochem5. This icon in a figure caption indicates a Living Figure that allows you to interact with three-dimensional representations of the illustration. Go to the Website and select the chapter and figure number.

The existence of RNA molecules that possess specific binding and catalytic properties makes plausible the idea of an early *"RNA world"* inhabited by life forms dependent on RNA molecules to play all major roles, including those important in heredity, the storage of information, and the promotion of specific reactions—that is, biosynthesis and energy metabolism.

2.2.3. Amino Acids and Their Polymers Can Play Biosynthetic and Catalytic Roles

In the early RNA world, the increasing populations of replicating RNA molecules would have consumed the building blocks of RNA that had been generated over long periods of time by prebiotic reactions. A shortage of these compounds would have favored the evolution of alternative mechanisms for their synthesis. A large number of pathways are possible. Examining the biosynthetic routes utilized by modern organisms can be a source of insight into which pathways survived. A striking observation is that simple amino acids are used as building blocks for the RNA bases (Figure 2.6). For both purines (adenine and guanine) and pyrimidines (uracil and cytosine), an amino acid serves as a core onto which the remainder of the base is elaborated. In addition, nitrogen atoms are donated by the amino group of the amino acid aspartic acid and by the amide group of the glutamine side chain.

Amino acids are chemically more versatile than nucleic acids because their side chains carry a wider range of chemical functionality. Thus, amino acids or short polymers of amino acids linked by *peptide bonds*, called *polypeptides* (Figure 2.7), may have functioned as components of ribozymes to provide a specific reactivity. Furthermore, longer polypeptides are capable of spontaneously folding to form well-defined three-dimensional structures, dictated by the sequence of amino acids along their polypeptide chains. The ability of polypeptides to fold spontaneously into elaborate structures, which permit highly specific chemical interactions with other molecules, may have favored the expansion of their roles in the course of evolution and is crucial to their dominant position in modern organisms. Today, most biological catalysts (enzymes) are not nucleic acids but are instead large polypeptides called *proteins*.

2.2.4. RNA Template-Directed Polypeptide Synthesis Links the RNA and Protein

Worlds

Polypeptides would have played only a limited role early in the evolution of life because their structures are not suited to self-replication in the way that nucleic acid structures are. However, polypeptides could have been included in evolutionary processes indirectly. For example, if the properties of a particular polypeptide favored the survival and replication of a class of RNA molecules, then these RNA molecules could have evolved ribozyme activities that promoted the synthesis of that polypeptide. This method of producing polypeptides with specific amino acid sequences has several limitations. First, it seems likely that only relatively short specific polypeptides could have been produced in this manner. Second, it would have been difficult to accurately link the particular amino acids in the polypeptide in a reproducible manner. Finally, a different ribozyme would have been required for each polypeptide. A critical point in evolution was reached when an apparatus for polypeptide synthesis developed that allowed *the sequence of bases in an RNA molecule to directly dictate the sequence of amino acids in a polypeptide*. A code evolved that established a relation between a specific sequence of three bases in RNA and an amino acid. We now call this set of three-base combinations, each encoding an amino acid, the *genetic code*. A decoding, or *translation*, system exists today as the *ribosome* and associated factors that are responsible for essentially all polypeptide synthesis from RNA templates in modern organisms. The essence of this mode of polypeptide synthesis is illustrated in Figure 2.8.

An RNA molecule (*messenger RNA*, or *mRNA*), containing in its base sequence the information that specifies a particular protein, acts as a template to direct the synthesis of the polypeptide. Each amino acid is brought to the template attached to an adapter molecule specific to that amino acid. These adapters are specialized RNA molecules (called *transfer RNAs* or *tRNAs*). After initiation of the polypeptide chain, a tRNA molecule with its associated amino acid binds to the template through specific Watson-Crick base-pairing interactions. Two such molecules bind to the ribosome and peptide-bond formation is catalyzed by an RNA component (called *ribosomal RNA* or *rRNA*) of the ribosome. The first RNA departs (with neither the polypeptide chain nor an amino acid attached) and another tRNA with its associated amino acid bonds to the ribosome. The growing polypeptide chain is transferred to this newly bound amino acid with the formation of a new peptide bond. This cycle then repeats itself. This scheme allows the sequence of the RNA template to encode the sequence of the polypeptide and thereby makes possible the production of long polypeptides with specified sequences. The mechanism of protein synthesis will be discussed in Chapter 29. Importantly, the ribosome is composed largely of RNA and is a highly sophisticated ribozyme, suggesting that it might be a surviving relic of the RNA world.

2.2.5. The Genetic Code Elucidates the Mechanisms of Evolution

The sequence of bases that encodes a functional protein molecule is called a *gene*. The genetic code—that is, the relation between the base sequence of a gene and the amino acid sequence of the polypeptide whose synthesis the gene directs—applies to all modern organisms with only very minor exceptions. This universality reveals that the genetic code was fixed early in the course of evolution and has been maintained to the present day.

We can now examine the mechanisms of evolution. Earlier, we considered how variation is required for evolution. We can now see that such variations in living systems are changes that alter the meaning of the genetic message. These variations are called *mutations*. A mutation can be as simple as a change in a single nucleotide (called a *point mutation*), such that a sequence of bases that encoded a particular amino acid may now encode another (Figure 2.9A). A mutation can also be the insertion or deletion of several nucleotides.

Other types of alteration permit the more rapid evolution of new biochemical activities. For instance, entire sections of the coding material can be duplicated, a process called *gene duplication* (Figure 2.9B). One of the duplication products may accumulate mutations and eventually evolve into a gene with a different, but related, function. Furthermore, parts of a gene may be duplicated and added to parts of another to give rise to a completely new gene, which encodes a protein with properties associated with each parent gene. Higher organisms contain many large families of enzymes and other macromolecules that are clearly related to one another in the same manner. Thus, gene duplication followed by specialization has been a crucial process in evolution. It allows the generation of macromolecules having particular functions without the need to start from scratch. The accumulation of genes with subtle and large differences allows for the generation of more complex biochemical processes and pathways and thus more complex organisms.

2.2.6. Transfer RNAs Illustrate Evolution by Gene Duplication

Transfer RNA molecules are the adaptors that associate an amino acid with its correct base sequence. Transfer RNA molecules are structurally similar to one another: each adopts a three-dimensional cloverleaf pattern of base-paired groups (Figure 2.10). Subtle differences in structure enable the protein-synthesis machinery to distinguish transfer RNA molecules with different amino acid specificities.

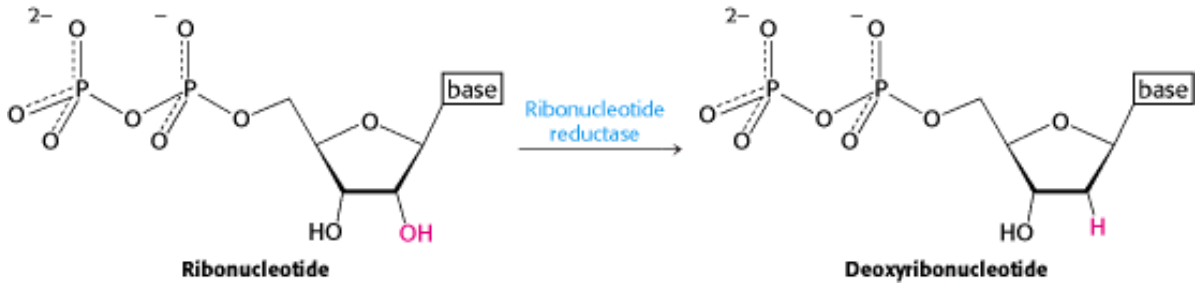
This family of related RNA molecules likely was generated by gene duplication followed by specialization. A nucleic acid sequence encoding one member of the family was duplicated, and the two copies evolved independently to generate molecules with specificities for different amino acids. This process was repeated, starting from one primordial transfer RNA gene until the 20 (or more) distinct members of the transfer RNA family present in modern organisms arose.

2.2.7. DNA Is a Stable Storage Form for Genetic Information

It is plausible that RNA was utilized to store genetic information early in the history of life. However, in modern organisms (with the exception of some viruses), the RNA derivative DNA (*deoxy*ribonucleic acid) performs this function (Sections 1.1.1 and 1.1.3). The 2′-hydroxyl group in the ribose unit of the RNA backbone is replaced by a hydrogen atom in DNA (Figure 2.11).

What is the selective advantage of DNA over RNA as the genetic material? The genetic material must be extremely stable so that sequence information can be passed on from generation to generation without degradation. RNA itself is a remarkably stable molecule; negative charges in the sugar-phosphate backbone protect it from attack by hydroxide ions that would lead to hydrolytic cleavage. However, the 2′-hydroxyl group makes the RNA susceptible to base-catalyzed hydrolysis. The removal of the 2′-hydroxyl group from the ribose decreases the rate of hydrolysis by approximately 100-fold under neutral conditions and perhaps even more under extreme conditions. Thus, the conversion of the genetic material from RNA into DNA would have substantially increased its chemical stability.

The evolutionary transition from RNA to DNA is recapitulated in the biosynthesis of DNA in modern organisms. In all cases, the building blocks used in the synthesis of DNA are synthesized from the corresponding building blocks of RNA by the action of enzymes termed *ribonucleotide reductases*. These enzymes convert ribonucleotides (a base and phosphate groups linked to a *ribose* sugar) into deoxyribonucleotides (a base and phosphates linked to *deoxyribose* sugar).

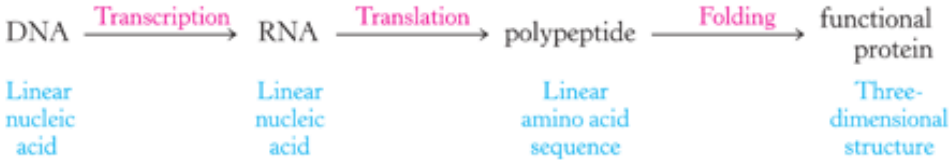


The properties of the ribonucleotide reductases vary substantially from species to species, but evidence suggests that they have a common mechanism of action and appear to have evolved from a common primordial enzyme.

The covalent structures of RNA and DNA differ in one other way. Whereas RNA contains *uracil*, DNA contains a methylated uracil derivative termed *thymine*. This modification also serves to protect the integrity of the genetic sequence, although it does so in a less direct manner. As we will see in Chapter 27, the methyl group present in thymine facilitates the repair of damaged DNA, providing an additional selective advantage.

Although DNA replaced RNA in the role of storing the genetic information, RNA maintained many of its other

functions. RNA still provides the template that directs polypeptide synthesis, the adaptor molecules, the catalytic activity of the ribosomes, and other functions. Thus, the genetic message is *transcribed* from DNA into RNA and then *translated* into protein.



This flow of sequence information from DNA to RNA to protein (to be considered in detail in Chapters 5, 28, and 29) applies to all modern organisms (with minor exceptions for certain viruses).

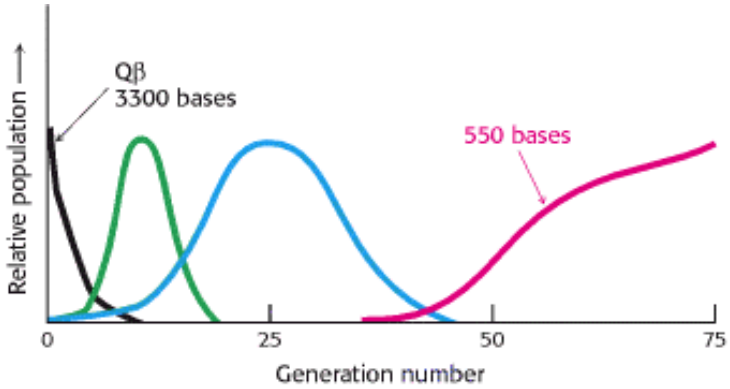
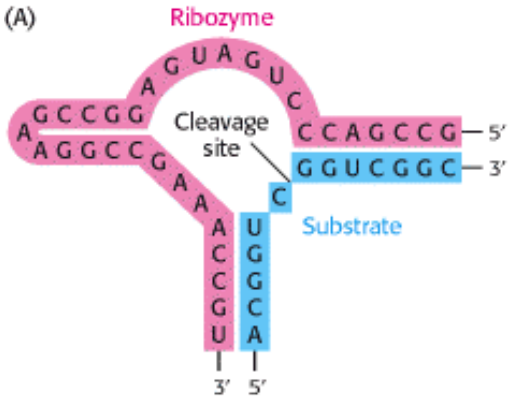


Figure 2.4. Evolution in a Test Tube. Rapidly replicating species of RNA molecules were generated from Q β RNA by exerting selective pressure. The green and blue curves correspond to species of intermediate size that accumulated and then became extinct in the course of the experiment.



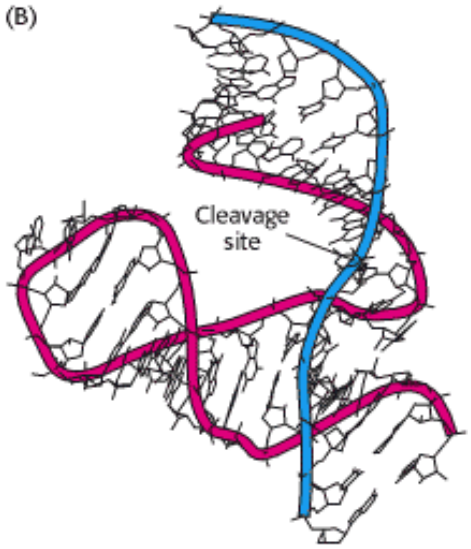


Figure 2.5. Catalytic RNA. (A) The base-pairing pattern of a "hammerhead" ribozyme and its substrate. (B) The folded conformation of the complex. The ribozyme cleaves the bond at the cleavage site. The paths of the nucleic acid backbones are highlighted in red and blue.

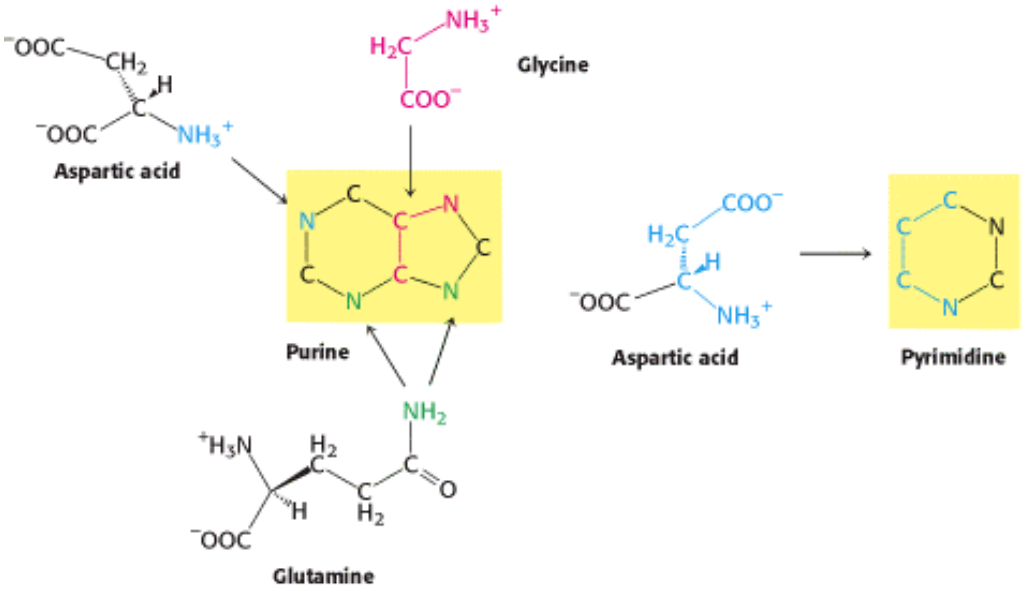


Figure 2.6. Biosynthesis of RNA Bases. Amino acids are building blocks for the biosynthesis of purines and pyrimidines.

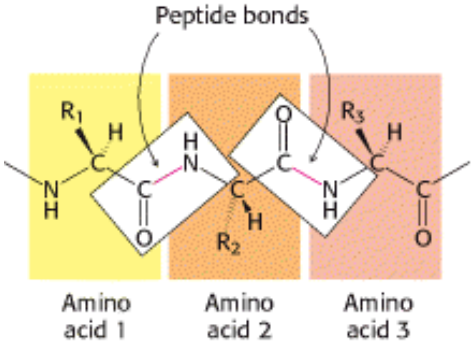


Figure 2.7. An Alternative Functional Polymer. Proteins are built of amino acids linked by peptide bonds.

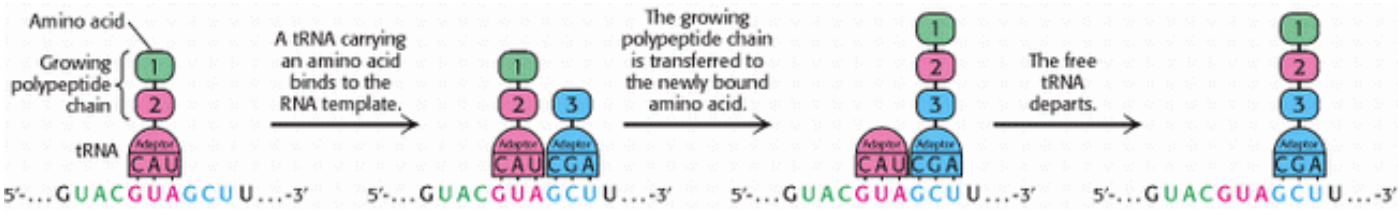


Figure 2.8. Linking the RNA and Protein Worlds. Polypeptide synthesis is directed by an RNA template. Adaptor RNA molecules, with amino acids attached, sequentially bind to the template RNA to facilitate the formation of a peptide bond between two amino acids. The growing polypeptide chain remains attached to an adaptor RNA until the completion of synthesis.

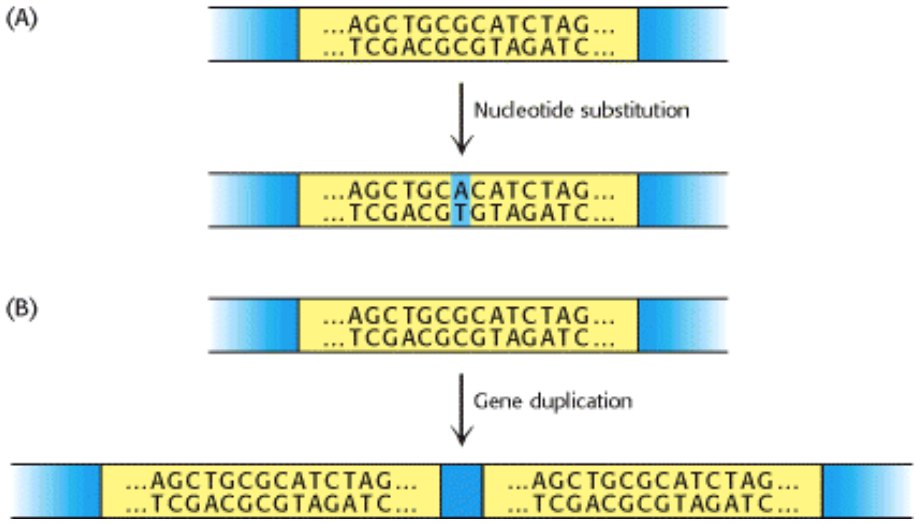


Figure 2.9. Mechanisms of Evolution. A change in a gene can be (A) as simple as a single base change or (B) as dramatic as partial or complete gene duplication.

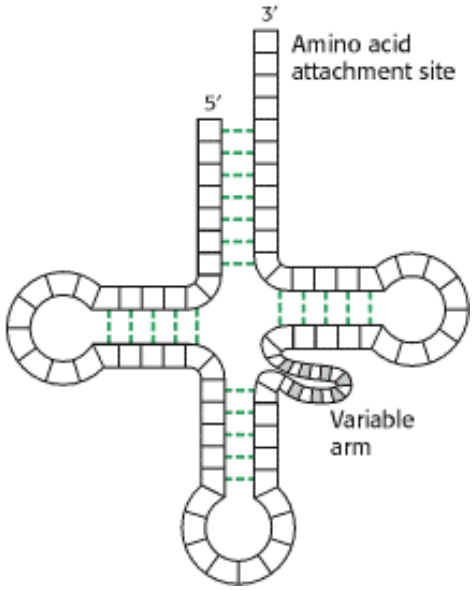


Figure 2.10. Cloverleaf Pattern of tRNA. The pattern of base-pairing interactions observed for all transfer RNA molecules reveals that these molecules had a common evolutionary origin.

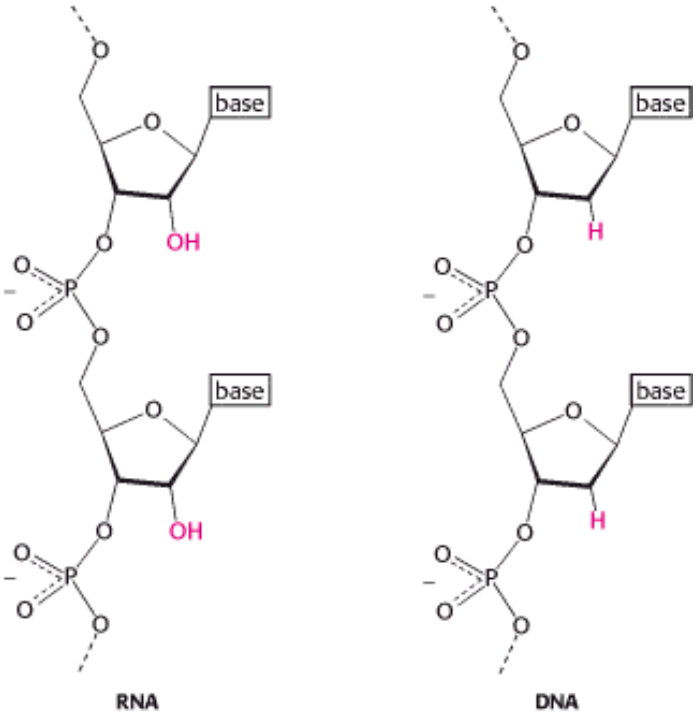


Figure 2.11. RNA and DNA Compared. Removal of the 2′-hydroxyl group from RNA to form DNA results in a backbone that is less susceptible to cleavage by hydrolysis and thus enables more-stable storage of genetic information.

2.3. Energy Transformations Are Necessary to Sustain Living Systems

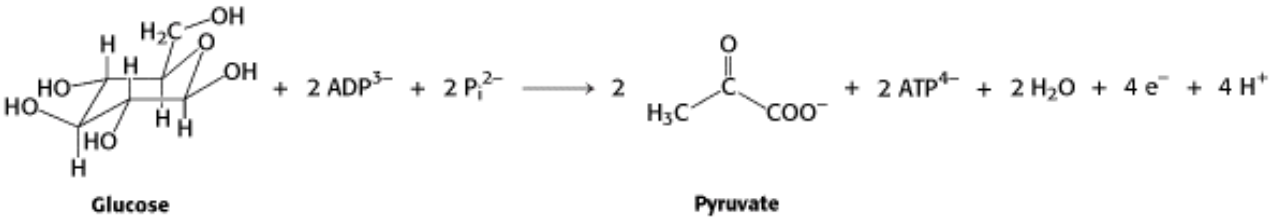
Most of the reactions that lead to the biosynthesis of nucleic acids and other biomolecules are not thermodynamically favorable under most conditions; they require an input of energy to proceed. Thus, they can proceed only if they are coupled to processes that release energy. How can energyrequiring and energy-releasing reactions be linked? How is energy from the environment transformed into a form that living systems can use? Answering these questions fundamental to biochemistry is the objective of much of this book.

2.3.1. ATP, a Common Currency for Biochemical Energy, Can Be Generated Through the Breakdown of Organic Molecules

Just as most economies simplify trade by using currency rather than bartering, biochemical systems have evolved common currencies for the exchange of energy. The most important of these currencies are molecules related to *adenosine triphosphate (ATP)* that contain an array of three linked phosphates (Figure 2.12). The bonds linking the phosphates persist in solution under a variety of conditions, but, when they are broken, an unusually large amount of energy is released that can be used to promote other processes. The roles of ATP and its use in driving other processes will be presented in detail in Chapter 14 and within many other chapters throughout this book.

ATP must be generated in appropriate quantities to be available for such reactions. The energy necessary for the synthesis of ATP can be obtained by the breakdown of other chemicals. Specific enzymes have evolved to couple these degradative processes to the phosphorylation of adenosine diphosphate (ADP) to yield ATP. Amino acids such as glycine, which were probably present in relatively large quantities in the prebiotic world and early in evolution, were likely sources of energy for ATP generation. The degradation of glycine to acetic acid may be an ATP-generation system that functioned early in evolution (Figure 2.13). In this reaction, the carbon-nitrogen bond in glycine is cleaved by reduction (the addition of electrons), and the energy released from the cleavage of this bond drives the coupling of ADP and orthophosphate (P_i) to produce ATP.

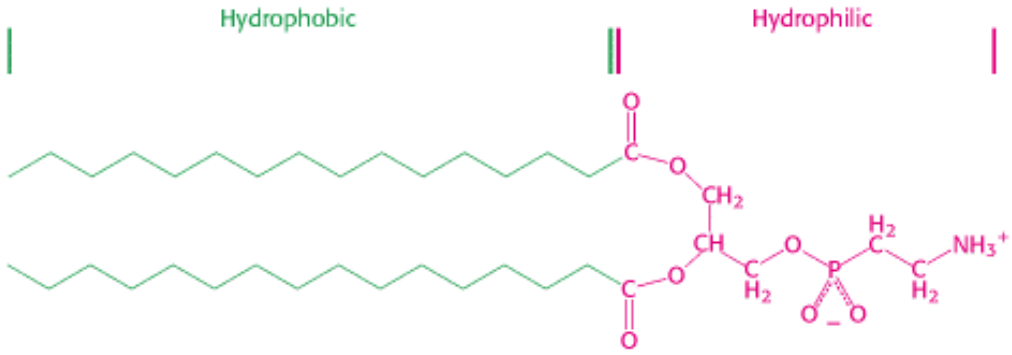
Amino acids are still broken down to produce ATP in modern organisms. However, sugars such as glucose are a more commonly utilized energy source because they are more readily metabolized and can be stored. The most important process for the direct synthesis of ATP in modern organisms is *glycolysis*, a complex process that derives energy from glucose.



Glycolysis presumably evolved as a process for ATP generation after carbohydrates such as glucose were being produced in significant quantities by other pathways. Glycolysis will be discussed in detail in Chapter 16.

2.3.2. Cells Were Formed by the Inclusion of Nucleic Acids Within Membranes

Modern organisms are made up of *cells*. A cell is composed of nucleic acids, proteins, and other biochemicals surrounded by a *membrane* built from lipids. These membranes completely enclose their contents, and so cells have a defined inside and outside. A typical membrane-forming lipid is phosphatidyl choline.



The most important feature of membrane-forming molecules such as phosphatidyl choline is that they are *amphipathic*—that is, they contain both *hydrophilic* (water-loving) and *hydrophobic* (water-avoiding) components. Membrane-forming molecules consist of fatty acids, whose long alkyl groups are hydrophobic, connected to shorter hydrophilic "head groups." When such lipids are in contact with water, they spontaneously aggregate to form specific structures such that the hydrophobic parts of the molecules are packed together away from water, whereas the hydrophilic parts are exposed to the aqueous solution. The structure that is important for membrane formation is the *lipid bilayer* (Figure 2.14). A bilayer is formed from two layers of lipids arranged such that the fatty acid tails of each layer interact with each other to form a hydrophobic interior while the hydrophilic head groups interact with the aqueous solution on each side. Such bilayer structures can fold onto themselves to form hollow spheres having interior compartments filled with water. The hydrophobic interior of the bilayer serves as a barrier between two aqueous phases. If such structures are formed in the presence of other molecules such as nucleic acids and proteins, these molecules can become trapped inside, thus forming cell-like structures. The structures of lipids and lipid bilayers will be considered in detail in Chapter 12.

At some stage in evolution, sufficient quantities of appropriate amphipathic molecules must have accumulated from biosynthetic or other processes to allow some nucleic acids to become entrapped and cell-like organisms to form. Such compartmentalization has many advantages. When the components of a cell are enclosed in a membrane, the products of enzymatic reactions do not simply diffuse away into the environment but instead are contained where they can be used by the cell that produced them. The containment is aided by the fact that nearly all biosynthetic intermediates and other biochemicals include one or more charged groups such as phosphates or carboxylates. Unlike more nonpolar or neutral

molecules, charged molecules do not readily pass through lipid membranes.

2.3.3. Compartmentalization Required the Development of Ion Pumps

Despite its many advantages, the enclosure of nucleic acids and proteins within membranes introduced several complications. Perhaps the most significant were the effects of *osmosis*. Membranes are somewhat permeable to water and small nonpolar molecules, whereas they are impermeable to macromolecules such as nucleic acids. When macromolecules are concentrated inside a compartment surrounded by such a semipermeable membrane, osmotic forces drive water through the membrane into the compartment. Without counterbalancing effects, the flow of water will burst the cell (Figure 2.15).

Osmosis-

The movement of a solvent across a membrane in the direction that tends to equalize concentrations of solute on the two sides of the membrane.

Modern cells have two distinct mechanisms for resisting these osmotic forces. One mechanism is to toughen the cell membrane by the introduction of an additional structure such as a cell wall. However, such a chemically elaborate structure may not have evolved quickly, especially because it must completely surround a cell to be effective. The other mechanism is the use of *energy-dependent ion pumps*. These pumps can lower the concentration of ions inside a cell relative to the outside, favoring the flow of water molecules from inside to outside. The resulting unequal distribution of ions across an inherently impermeable membrane is called an *ion gradient*. Appropriate ion gradients can balance the osmotic forces and maintain a cell at a constant volume. Membrane proteins such as ion pumps will be considered in Chapter 13.

Ion gradients can prevent osmotic crises, but they require energy to be produced. Most likely, an ATP-driven proton pump was the first existing component of the machinery for generating an ion gradient (Figure 2.16). Such pumps, which are found in essentially all modern cells, hydrolyze ATP to ADP and inorganic phosphate and utilize the energy released to transport protons from the inside to the outside of a cell. The pump thus establishes a proton gradient that, in turn, can be coupled to other membrane-transport processes such as the removal of sodium ions from the cell. The proton gradient and other ion gradients generated from it act together to counteract osmotic effects and prevent the cell from swelling and bursting.

2.3.4. Proton Gradients Can Be Used to Drive the Synthesis of ATP

Enzymes act to accelerate reactions, but they cannot alter the position of chemical equilibria. An enzyme that accelerates a reaction in the forward direction must also accelerate the reaction to the same extent in the reverse direction. Thus, the existence of an enzyme that utilized the hydrolysis of ATP to generate a proton gradient presented a tremendous opportunity for the evolution of alternative systems for generating ATP. Such an enzyme could synthesize ATP by reversing the process that produces the gradient. Enzymes, now called *ATP synthases*, do in fact use proton gradients to drive the bonding of ADP and P_i to form ATP (Figure 2.17). These proteins will be considered in detail in Chapter 18.

Organisms have evolved a number of elaborate mechanisms for the generation of proton gradients across membranes. An example is *photosynthesis*, a process first used by bacteria and now also used by plants to harness the light energy from the sun. The essence of photosynthesis is the light-driven transfer of an electron across a membrane. The fundamental processes are illustrated in Figure 2.18.

The photosynthetic apparatus, which is embedded in a membrane, contains pigments that efficiently absorb light from the sun. The absorbed light provides the energy to promote an electron in the pigment molecule to an excited state. The

high-energy electron can then jump to an appropriate acceptor molecule located in the part of the membrane facing the inside of the cell. The acceptor molecule, now reduced, binds a proton from a water molecule, generating an hydroxide ion inside the cell. The electronic "hole" left in the pigment on the outside of the membrane can then be filled by the donation of an electron from a suitable reductant on the outside of the membrane. Because the generation of an hydroxide ion inside the cell is equivalent to the generation of a proton outside the cell, a proton gradient develops across the membrane. Protons flow down this gradient through ATP synthases to generate ATP.

Photosynthesis is but one of a range of processes in different organisms that lead to ATP synthesis through the action of proteins evolutionarily related to the primordial ATP-driven pumps. In animals, the degradation of carbohydrates and other organic compounds is the source of the electron flow across membranes that can be used to develop proton gradients. The formation of ATP-generating proton gradients by fuel metabolism will be considered in Chapter 18 and by light absorption in Chapter 19.

2.3.5. Molecular Oxygen, a Toxic By-Product of Some Photosynthetic Processes, Can Be Utilized for Metabolic Purposes

As stated earlier, photosynthesis generates electronic "holes" in the photosynthetic apparatus on the outside of the membrane. These holes are powerful oxidizing agents; that is, they have very high affinities for electrons and can pull electrons from many types of molecules. They can even oxidize water. Thus, for many photosynthetic organisms, the electron donor that completes the photosynthetic cycle is water. The product of water oxidation is oxygen gas—that is, molecular oxygen (O_2).

$$2\ H_2O \longrightarrow O_2 + 4\ e^- + 4\ H^+$$

The use of water as the electron donor significantly increases the efficiency of photosynthetic ATP synthesis because the generation of one molecule of oxygen is accompanied not only by the release of four electrons (e^-), but also by the release of four protons on one side of the membrane. Thus, an additional proton is released for each proton equivalent produced by the initial electron-transfer process, so twice as many protons are available to drive ATP synthesis. Oxygen generation will be considered in Chapter 19.

Oxygen was present in only small amounts in the atmosphere before organisms evolved that could oxidize water. The "pollution" of the air with oxygen produced by photosynthetic organisms greatly affected the course of evolution. Oxygen is quite reactive and thus extremely toxic to many organisms. Many biochemical processes have evolved to protect cells from the deleterious effects of oxygen and other reactive species that can be generated from this molecule. Subsequently, organisms evolved mechanisms for taking advantage of the high reactivity of oxygen to promote favorable processes. Most important among these mechanisms are those for the oxidation of organic compounds such as glucose. Through the action of oxygen, a glucose molecule can be completely converted into carbon dioxide and water, releasing enough energy to synthesize approximately 30 molecules of ATP.

$$\text{Glucose} + 6\ O_2 \longrightarrow 6\ CO_2 + 6\ H_2O + \text{energy}$$

This number represents a 15-fold increase in ATP yield compared with the yield from the breakdown of glucose in the absence of oxygen in the process of glycolysis. This increased efficiency is apparent in everyday life; our muscles exhaust their fuel supply and tire quickly if they do not receive enough oxygen and are forced to use glycolysis as the sole ATP source. The role of oxygen in the extraction of energy from organic molecules will be considered in Chapter 18.

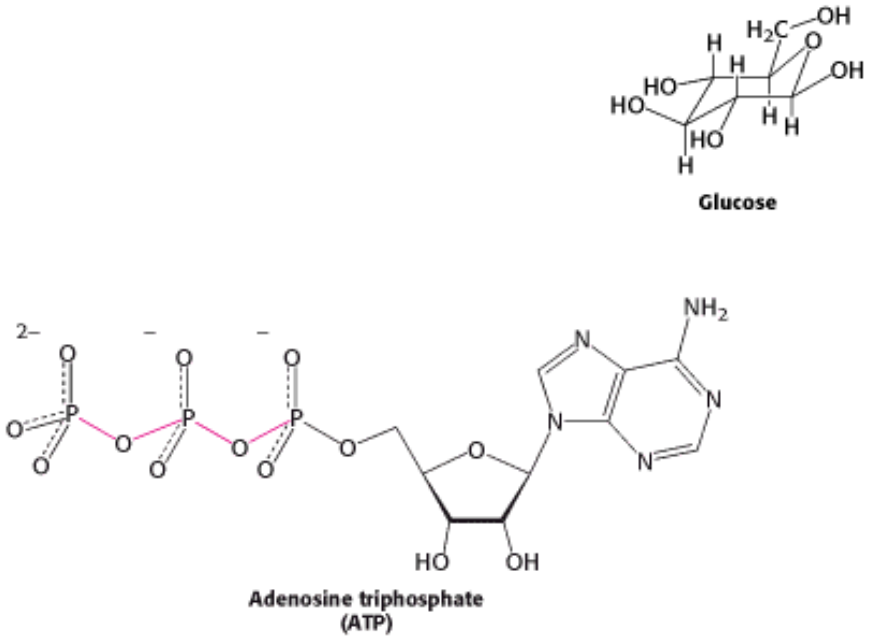


Figure 2.12. ATP, the Energy Currency of Living Systems. The phosphodiester bonds (red) release considerable energy when cleaved by hydrolysis or other processes.

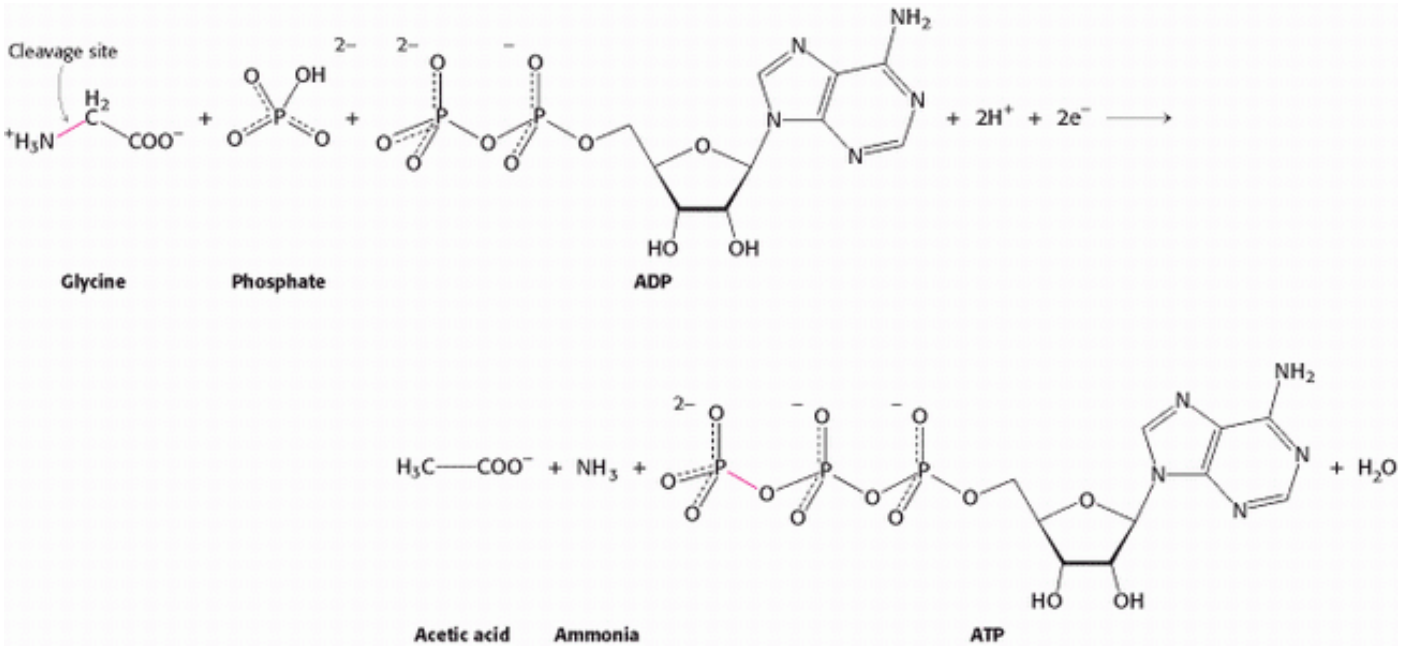


Figure 2.13. A Possible Early Method for Generating ATP. The synthesis of ATP might have been driven by the degradation of glycine.

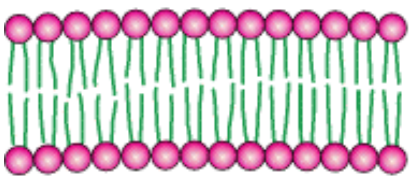


Figure 2.14. Schematic View of a Lipid Bilayer. These structures define the boundaries of cells.

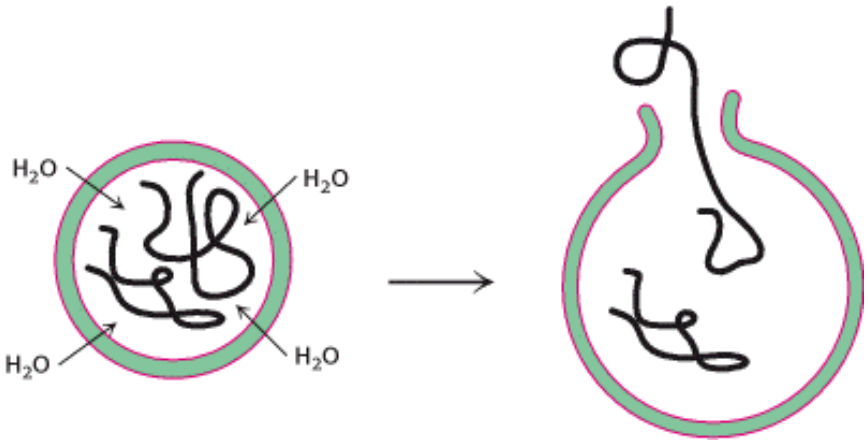


Figure 2.15. The "Osmotic Crisis." A cell consisting of macromolecules surrounded by a semipermeable membrane will take up water from outside the cell and burst.

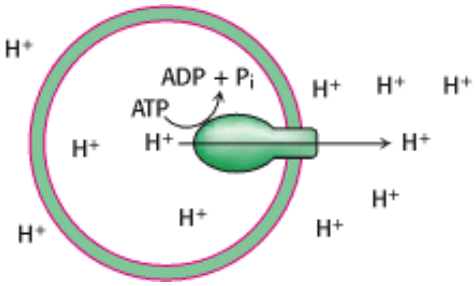


Figure 2.16. Generating an Ion Gradient. ATP hydrolysis can be used to drive the pumping of protons (or other ions) across a membrane.

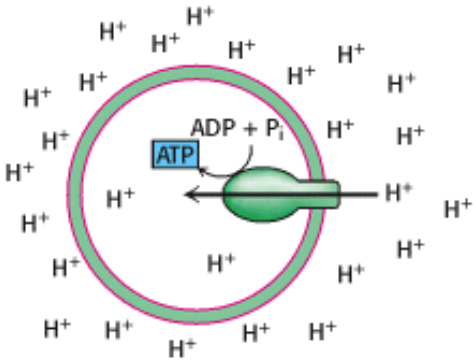


Figure 2.17. Use of Proton Gradients to Synthesize ATP. ATP can be synthesized by the action of an ATP-driven proton pump running in reverse.

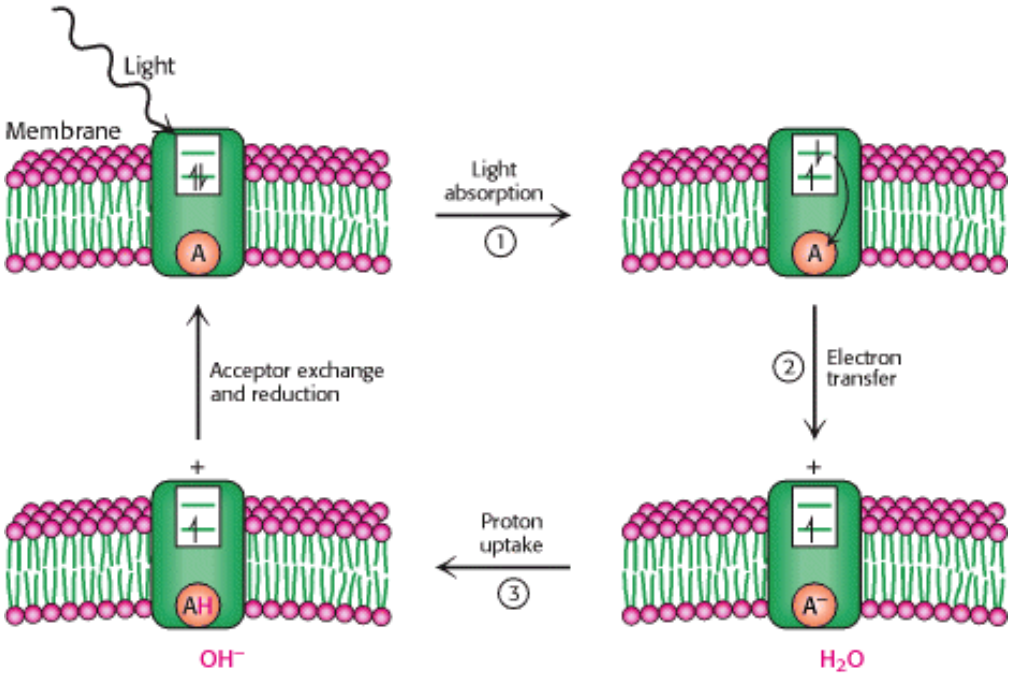


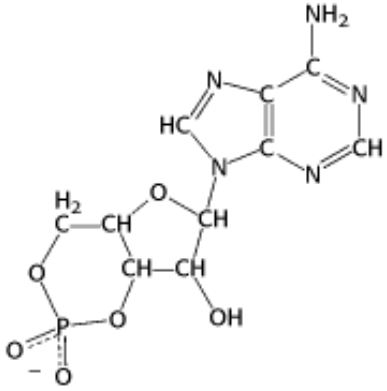
Figure 2.18. Photosynthesis. Absorption of light (1) leads to electron transfer across a membrane (2). For each electron transfer, one excess hydroxide ion is generated inside the cell (3). The process produces a proton gradient across the membrane that can drive ATP synthesis.

2.4. Cells Can Respond to Changes in Their Environments

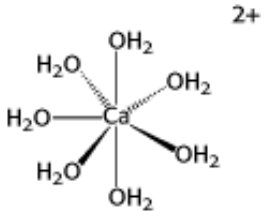
The environments in which cells grow often change rapidly. For example, cells may consume all of a particular food source and must utilize others. To survive in a changing world, cells evolved mechanisms for adjusting their biochemistry in response to signals indicating environmental change. The adjustments can take many forms, including changes in the activities of preexisting enzyme molecules, changes in the rates of synthesis of new enzyme molecules, and changes in membrane-transport processes.

Initially, the detection of environmental signals occurred inside cells. Chemicals that could pass into cells, either by diffusion through the cell membrane or by the action of transport proteins, and could bind directly to proteins inside the cell and modulate their activities. An example is the use of the sugar arabinose by the bacterium *Escherichia coli* (Figure 2.19). *E. coli* cells are normally unable to use arabinose efficiently as a source of energy. However, if arabinose is their only source of carbon, *E. coli* cells synthesize enzymes that catalyze the conversion of this sugar into useful forms. This response is mediated by arabinose itself. If present in sufficient quantity outside the cell, arabinose can enter the cell through transport proteins. Once inside the cell, arabinose binds to a protein called AraC. This binding alters the structure of AraC so that it can now bind to specific sites in the bacterial DNA and increase RNA transcription from genes encoding enzymes that metabolize arabinose. The mechanisms of gene regulation will be considered in Chapter 31.

Subsequently, mechanisms appeared for detecting signals at the cell surface. Cells could thus respond to signaling molecules even if those molecules did not pass into the cell. Receptor proteins evolved that, embedded in the membrane, could bind chemicals present in the cellular environment. Binding produced changes in the protein structure that could be detected at the inside surface of the cell membrane. By this means, chemicals outside the cell could influence events inside the cell. Many of these *signal-transduction pathways* make use of substances such as cyclic adenosine monophosphate (cAMP) and calcium ion as "second messengers" that can diffuse throughout the cell, spreading the word of environmental change.



Cyclic AMP (cAMP)



Calcium ion in water

The second messengers may bind to specific sensor proteins inside the cell and trigger responses such as the activation of enzymes. Signal-transduction mechanisms will be considered in detail in Chapter 15 and in many other chapters throughout this book.

2.4.1. Filamentous Structures and Molecular Motors Enable Intracellular and Cellular Movement

The development of the ability to move was another important stage in the evolution of cells capable of adapting to a changing environment. Without this ability, nonphotosynthetic cells might have starved after consuming the nutrients available in their immediate vicinity.

Bacteria swim through the use of filamentous structures termed *flagella* that extend from their cell membranes (Figure 2.20). Each bacterial cell has several flagella, which, under appropriate conditions, form rotating bundles that efficiently propel the cell through the water. These flagella are long polymers consisting primarily of thousands of identical protein subunits. At the base of each flagellum are assemblies of proteins that act as motors to drive its rotation. The rotation of the flagellar motor is driven by the flow of protons from outside to inside the cell. Thus, energy stored in the form of a proton gradient is transduced into another form, rotatory motion.

Other mechanisms for motion, also depending on filamentous structures, evolved in other cells. The most important of these structures are *microfilaments* and *microtubules*. Microfilaments are polymers of the protein *actin*, and microtubules are polymers of two closely related proteins termed α - and β-*tubulin*. Unlike a bacterial flagellum, these filamentous structures are highly dynamic: they can rapidly increase or decrease in length through the addition or subtraction of component protein molecules. Microfilaments and microtubules also serve as tracks on which other proteins move, driven by the hydrolysis of ATP. Cells can change shape through the motion of *molecular motor proteins* along such filamentous structures that are changing in shape as a result of dynamic polymerization (Figure 2.21). Coordinated shape changes can be a means of moving a cell across a surface and are crucial to cell division. The motor proteins are also responsible for the transport of organelles and other structures within eukaryotic cells. Molecular motors will be considered in Chapter 34.

2.4.2. Some Cells Can Interact to Form Colonies with Specialized Functions

Early organisms lived exclusively as single cells. Such organisms interacted with one another only indirectly by competing for resources in their environments. Certain of these organisms, however, developed the ability to form colonies comprising many interacting cells. In such groups, the environment of a cell is dominated by the presence of surrounding cells, which may be in direct contact with one another. These cells communicate with one another by a variety of signaling mechanisms and may respond to signals by altering enzyme activity or levels of gene expression. One result may be *cell differentiation;* differentiated cells are genetically identical but have different properties because their genes are expressed differently.

Several modern organisms are able to switch back and forth from existence as independent single cells to existence as multicellular colonies of differentiated cells. One of the most well characterized is the slime mold *Dictyostelium*. In favorable environments, this organism lives as individual cells; under conditions of starvation, however, the cells come together to form a cell aggregate. This aggregate, sometimes called a *slug*, can move as a unit to a potentially more favorable environment where it then forms a multicellular structure, termed a *fruiting body*, that rises substantially above the surface on which the cells are growing. Wind may carry cells released from the top of the fruiting body to sites where the food supply is more plentiful. On arriving in a well-stocked location, the cells grow, reproduce, and live as individual cells until the food supply is again exhausted (Figure 2.22).

The transition from unicellular to multicellular growth is triggered by cell-cell communication and reveals much about signaling processes between and within cells. Under starvation conditions, *Dictyostelium* cells release the signal molecule cyclic AMP. This molecule signals surrounding cells by binding to a membrane-bound protein receptor on the cell surface. The binding of cAMP molecules to these receptors triggers several responses, including movement in the direction of higher cAMP concentration, as well as the generation and release of additional cAMP molecules (Figure 2.23).

The cells aggregate by following cAMP gradients. Once in contact, they exchange additional signals and then differentiate into distinct *cell types*, each of which expresses the set of genes appropriate for its eventual role in forming the fruiting body (Figure 2.24). The life cycles of organisms such as *Dictyostelium* foreshadow the evolution of organisms that are multicellular throughout their lifetimes. It is also interesting to note the cAMP signals starvation in many organisms, including human beings.

2.4.3. The Development of Multicellular Organisms Requires the Orchestrated Differentiation of Cells

The fossil record indicates that macroscopic, multicellular organisms appeared approximately 600 million years ago. Most of the organisms familiar to us consist of many cells. For example, an adult human being contains approximately 100,000,000,000,000 cells. The cells that make up different organs are distinct and, even within one organ, many different cell types are present. Nonetheless, the DNA sequence in each cell is identical. The differences between cell types are the result of differences in how these genes are expressed.

Each multicellular organism begins as a single cell. For this cell to develop into a complex organism, the embryonic cells must follow an intricate program of regulated gene expression, cell division, and cell movement. The developmental program relies substantially on the responses of cells to the environment created by neighboring cells. Cells in specific positions within the developing embryo divide to form particular tissues, such as muscle. Developmental pathways have been extensively studied in a number of organisms, including the nematode *Caenorhabditis elegans* (Figure 2.25), a 1-mm-long worm containing 959 cells. A detailed map describing the fate of each cell in *C. elegans* from the fertilized egg to the adult is shown in Figure 2.26. Interestingly, proper development requires not only cell division but also the death of specific cells at particular points in time through a process called programmed cell death or *apoptosis*.

Investigations of genes and proteins that control development in a wide range of organisms have revealed a great many common features. Many of the molecules that control human development are evolutionarily related to those in relatively simple organisms such as *C. elegans*. Thus, solutions to the problem of controlling development in multicellular organisms arose early in evolution and have been adapted many times in the course of evolution, generating the great diversity of complex organisms.

2.4.4. The Unity of Biochemistry Allows Human Biology to Be Effectively Probed Through Studies of Other Organisms

All organisms on Earth have a common origin (Figure 2.27). How could complex organisms such as human beings have evolved from the simple organisms that existed at life's start? The path outlined in this chapter reveals that most of the fundamental processes of biochemistry were largely fixed early in the history of life. The complexity of organisms such

as human beings is manifest, at a biochemical level, in the interactions between overlapping and competing pathways, which lead to the generation of intricately connected groups of specialized cells. The evolution of biochemical and physiological complexity is made possible by the effects of gene duplication followed by specialization. Paradoxically, the reliance on gene duplication also makes this complexity easier to comprehend. Consider, for example, the protein kinases—enzymes that transfer phosphoryl groups from ATP to specific amino acids in proteins. These enzymes play essential roles in many signal-transduction pathways and in the control of cell growth and differentiation. The human genome encodes approximately 500 proteins of this class; even a relatively simple, unicellular organism such as brewer's yeast has more than 100 protein kinases. Yet each of these enzymes is the evolutionary descendant of a common ancestral enzyme. Thus, *we can learn much about the essential behavior of this large collection of proteins through studies of a single family member*. After the essential behavior is understood, we can evaluate the specific adaptations that allow each family member to perform its particular biological functions.

Most central processes in biology have been characterized first in relatively simple organisms, often through a combination of genetic, physiological, and biochemical studies. Many of the processes controlling early embryonic development were elucidated by the results of studies of the fruit fly. The events controlling DNA replication and the cell cycle were first deciphered in yeast. Investigators can now test the functions of particular proteins in mammals by disrupting the genes that encode these proteins in mice and examining the effects. The investigations of organisms linked to us by common evolutionary pathways are powerful tools for exploring all of biology and for developing new understanding of normal human function and disease.

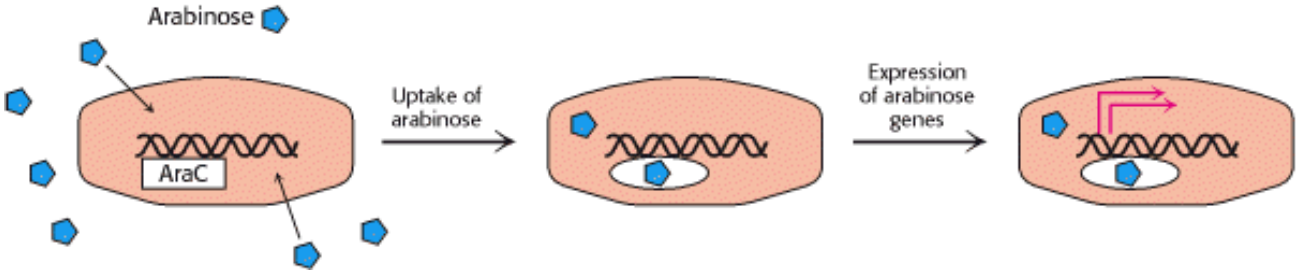


Figure 2.19. Responding to Environmental Conditions. In *E. coli* cells, the uptake of arabinose from the environment triggers the production of enzymes necessary for its utilization.

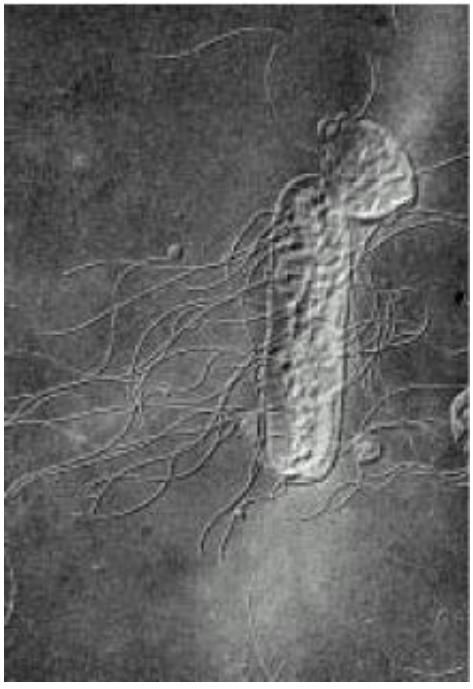


Figure 2.20. Bacteria with Flagella. A bacterium (*Proteus mirabilis*) swims through the rotation of filamentous structures called flagella. [Fred E. Hossler/ Visuals Unlimited.]

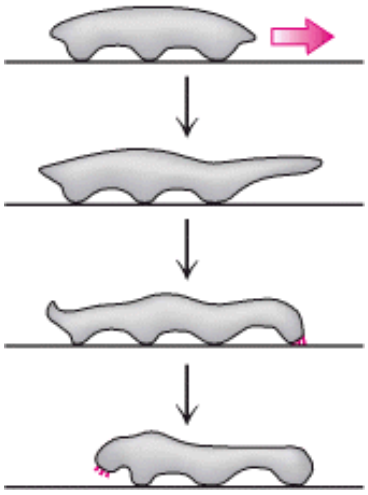


Figure 2.21. Alternative Movement. Cell mobility can be achieved by changes in cell shape.

Figure 2.22. Unicellular to Multicellular Transition in *Dictyostelium*. This scanning electron migrograph shows the transformation undergone by the slime mold *Dictyostelium*. Hundreds of thousands of single cells aggregate to form a migrating slug, seen in the lower left. Once the slug comes to a stop, it gradually elongates to form the fruiting body. [Courtesy of M. J. Grimsom and R. L. Blanton, Texas Tech University.]

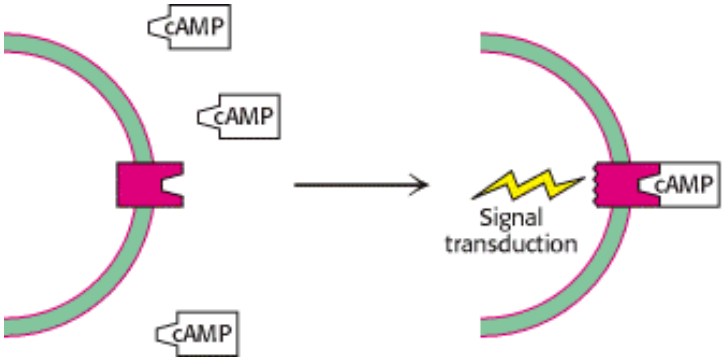


Figure 2.23. Intracellular Signaling. Cyclic AMP, detected by cell-surface receptors, initiates the formation of aggregates in *Dictyostelium*.

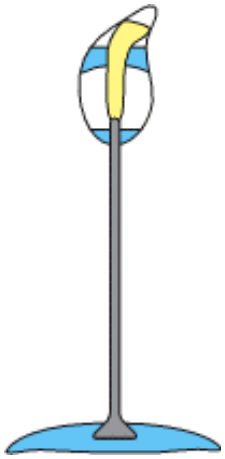


Figure 2.24. Cell Differentiation in *Dictyostelium*. The colors represent the distribution of cell types expressing similar sets of genes in the *Dictyostelium* fruiting body.

Figure 2.25. The Nematode *Caenorhabditis elegans*. This organism serves as a useful model for development. [Sinclair Stammers Science Photo Library/Photo Researchers.]

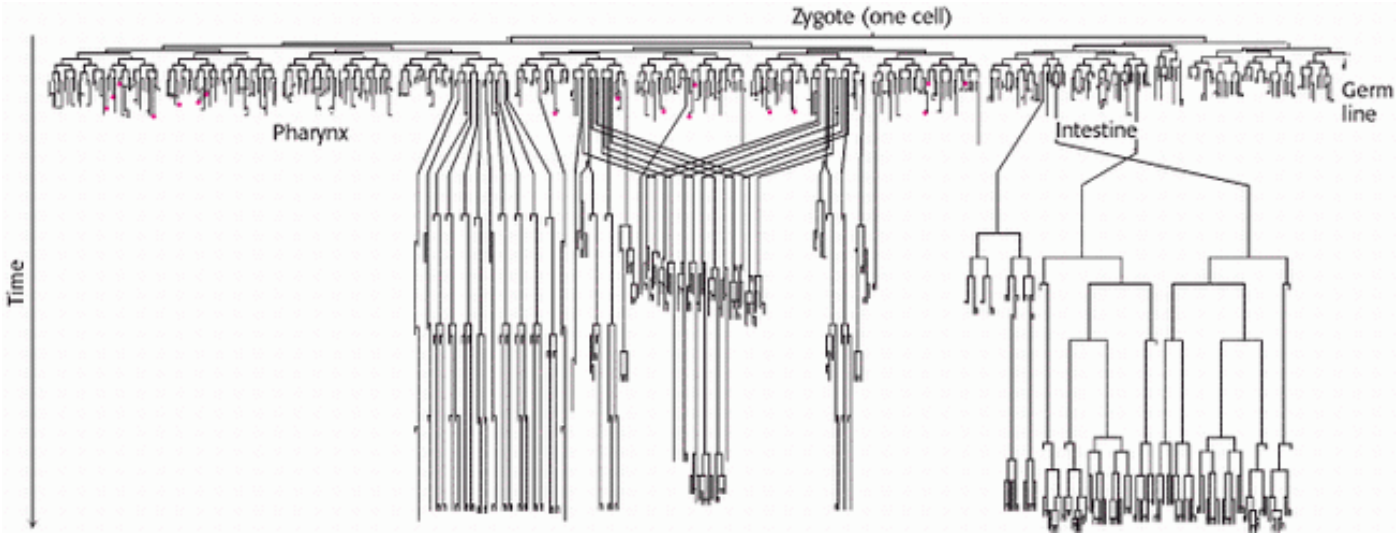


Figure 2.26. Developmental Pathways of *C. elegans*. The nematode develops from a single cell, called a zygote, into a complex organism. The fate of each individual cell in *C. elegans* is known and can be followed by referring to the cell-lineage diagram. The labels indicate cells that form specific organs. Cells that undergo programmed cell death are shown in red.

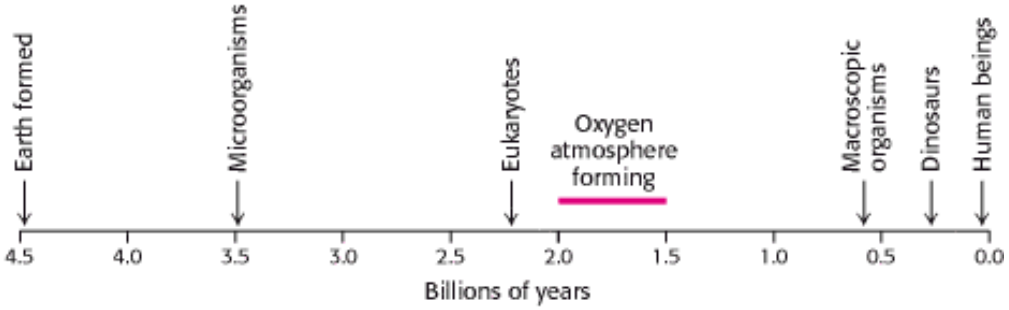


Figure 2.27. A Possible Time Line for Biochemical Evolution. Key events are indicated.

Summary

Key Organic Molecules Are Used by Living Systems

The evolution of life required a series of transitions, beginning with the generation of organic molecules that could serve as the building blocks for complex biomolecules. How these molecules arose is a matter of conjecture, but experiments have established that they could have formed under hypothesized prebiotic conditions.

Evolution Requires Reproduction, Variation, and Selective Pressure

The next major transition in the evolution of life was the formation of replicating molecules. Replication, coupled with variation and selective pressure, marked the beginning of evolution. Variation was introduced by a number of means, from simple base substitutions to the duplication of entire genes. RNA appears to have been an early replicating molecule. Furthermore, some RNA molecules possess catalytic activity. However, the range of reactions that RNA is capable of catalyzing is limited. With time, the catalytic activity was transferred to proteins—linear polymers of the

chemically versatile amino acids. RNA directed the synthesis of these proteins and still does in modern organisms through the development of a genetic code, which relates base sequence to amino acid sequence. Eventually, RNA lost its role as the gene to the chemically similar but more stable nucleic acid DNA. In modern organisms, RNA still serves as the link between DNA and protein.

Energy Transformations Are Necessary to Sustain Living Systems

Another major transition in evolution was the ability to transform environmental energy into forms capable of being used by living systems. ATP serves as the cellular energy currency that links energy-yielding reactions with energy-requiring reactions. ATP itself is a product of the oxidation of fuel molecules, such as amino acids and sugars. With the evolution of membranes—hydrophobic barriers that delineate the borders of cells—ion gradients were required to prevent osmotic crises. These gradients were formed at the expense of ATP hydrolysis. Later, ion gradients generated by light or the oxidation of fuel molecules were used to synthesize ATP.

Cells Can Respond to Changes in Their Environments

The final transition was the evolution of sensing and signaling mechanisms that enabled a cell to respond to changes in its environment. These signaling mechanisms eventually led to cell-cell communication, which allowed the development of more-complex organisms. The record of much of what has occurred since the formation of primitive organisms is written in the genomes of extant organisms. Knowledge of these genomes and the mechanisms of evolution will enhance our understanding of the history of life on Earth as well as our understanding of existing organisms.

Key Terms

prebiotic world

reproduction

variation

competition

selective pressure

catalyst

enzyme

ribozyme

RNA world

proteins

genetic code

translation

gene

mutation

gene duplication

ATP (adenosine triphosphate)

membrane

ion pump

ion gradient

photosynthesis

signal transduction pathway

molecular motor protein

cell differentiation

unity of biochemistry

Problems

1. *Finding the fragments*. Identify the likely source (CH_4, NH_3, H_2O, or H_2) of each atom in alanine generated in the Miller-Urey experiment.

   See answer

2. *Following the populations*. In an experiment analogous to the Spiegelman experiment, suppose that a population of RNA molecules consists of 99 identical molecules, each of which replicates once in 15 minutes, and 1 molecule that replicates once in 5 minutes. Estimate the composition of the population after 1, 10, and 25 "generations" if a generation is defined as 15 minutes of replication. Assume that all necessary components are readily available.

   See answer

3. *Selective advantage*. Suppose that a replicating RNA molecule has a mutation (genotypic change) and the phenotypic result is that it binds nucleotide monomers more tightly than do other RNA molecules in its population. What might the selective advantage of this mutation be? Under what conditions would you expect this selective advantage to be most important?

   See answer

4. *Opposite of randomness*. Ion gradients prevent osmotic crises, but they require energy to be produced. Why does the formation of a gradient require an energy input?

   See answer

5. *Coupled gradients*. How could a proton gradient with a higher concentration of protons inside a cell be used to pump ions out of a cell?

   See answer

6. *Proton counting*. Consider the reactions that take place across a photosynthetic membrane. On one side of the membrane, the following reaction takes place:

$$4\,e^- + 4\,A^- + 4\,H_2O \longrightarrow 4\,AH + 4\,OH^-$$

Need extra help? Purchase chapters of the Student Companion with complete solutions online at www.whfreeman.com/ biochem5.

   whereas, on the other side of the membrane, the reaction is:

$$2\,H_2O \longrightarrow O_2 + 4\,e^- + 4\,H^+$$

   How many protons are made available to drive ATP synthesis for each reaction cycle?

   See answer

7. *An alternative pathway*. To respond to the availability of sugars such as arabinose, a cell must have at least two types of proteins: a transport protein to allow the arabinose to enter the cell and a gene-control protein, which binds the arabinose and modifies gene expression. To respond to the availability of some very hydrophobic molecules, a cell requires only one protein. Which one and why?

   See answer

8. *How many divisions?* In the development pathway of *C. elegans,* cell division is initially synchronous—that is, all cells divide at the same rate. Later in development, some cells divide more frequently than do others. How many times does each cell divide in the synchronous period? Refer to Figure 2.26.

   See answer

Selected Readings

Where to start

N.R. Pace. 2000. The universal nature of biochemistry *Proc. Natl. Acad. Sci. U. S. A.* 98: 805-808. (PubMed) (Full Text in PMC)

L.E. Orgel. 1987. Evolution of the genetic apparatus: A review *Cold Spring Harbor Symp. Quant. Biol.* 52: 9-16. (PubMed)

A. Lazcano and S.L. Miller. 1996. The origin and early evolution of life: Prebiotic chemistry, the pre-RNA world, and time *Cell* 85: 793-798. (PubMed)

L.E. Orgel. 1998. The origin of life: A review of facts and speculations *Trends Biochem. Sci.* 23: 491-495. (PubMed)

Books

Darwin, C., 1975. *On the Origin of Species, a Facsimile of the First Edition* . Harvard University Press.

Gesteland, R. F., Cech, T., and Atkins, J. F., 1999. *The RNA World* . Cold Spring Harbor Laboratory Press.

Dawkins, R., 1996. *The Blind Watchmaker* . Norton.

Smith, J. M., and Szathmáry, E., 1995. *The Major Transitions in Evolution* . W. H. Freeman and Company.

Prebiotic chemistry

S.L. Miller. 1987. Which organic compounds could have occurred on the prebiotic earth? *Cold Spring Harbor Symp. Quant. Biol.* 52: 17-27. (PubMed)

F.H. Westheimer. 1987. Why nature chose phosphates *Science* 235: 1173-1178. (PubMed)

M. Levy and S.L. Miller. 1998. The stability of the RNA bases: Implications for the origin of life *Proc. Natl. Acad. Sci. U. S. A.* 95: 7933-7938. (PubMed) (Full Text in PMC)

R. Sanchez, J. Ferris, and L.E. Orgel. 1966. Conditions for purine synthesis: Did prebiotic synthesis occur at low temperatures? *Science* 153: 72-73. (PubMed)

In vitro evolution

D.R. Mills, R.L. Peterson, and S. Spiegelman. 1967. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule *Proc. Natl. Acad. Sci. U. S. A.* 58: 217-224. (PubMed)

R. Levisohn and S. Spiegelman. 1969. Further extracellular Darwinian experiments with replicating RNA molecules: Diverse variants isolated under different selective conditions *Proc. Natl. Acad. Sci. U. S. A.* 63: 805-811. (PubMed)

D.S. Wilson and J.W. Szostak. 1999. In vitro selection of functional nucleic acids *Annu. Rev. Biochem.* 68: 611-647. (PubMed)

Replication and catalytic RNA

T.R. Cech. 1993. The efficiency and versatility of catalytic RNA: Implications for an RNA world *Gene* 135: 33-36. (PubMed)

L.E. Orgel. 1992. Molecular replication *Nature* 358: 203-209. (PubMed)

W.S. Zielinski and L.E. Orgel. 1987. Autocatalytic synthesis of a tetranucleotide analogue *Nature* 327: 346-347. (PubMed)

K.E. Nelson, M. Levy, and S.L. Miller. 2000. Peptide nucleic acids rather than RNA may have been the first genetic molecule *Proc. Natl. Acad. Sci. U. S. A.* 97: 3868-3871. (PubMed) (Full Text in PMC)

Transition from RNA to DNA

P. Reichard. 1997. The evolution of ribonucleotide reduction *Trends Biochem. Sci.* 22: 81-85. (PubMed)

A. Jordan and P. Reichard. 1998. Ribonucleotide reductases *Annu. Rev. Biochem.* 67: 71-98. (PubMed)

Membranes

T.H. Wilson and P.C. Maloney. 1976. Speculations on the evolution of ion transport mechanisms *Fed. Proc.* 35: 2174-2179. (PubMed)

T.H. Wilson and E.C. Lin. 1980. Evolution of membrane bioenergetics *J. Supramol. Struct.* 13: 421-446. (PubMed)

Multicellular organisms and development

G. Mangiarotti, S. Bozzaro, S. Landfear, and H.F. Lodish. 1983. Cell-cell contact, cyclic AMP, and gene expression during development of *Dictyostelium discoideum Curr. Top. Dev. Biol.* 18: 117-154. (PubMed)

C. Kenyon. 1988. The nematode *Caenorhabditis elegans Science* 240: 1448-1453. (PubMed)

J. Hodgkin, R.H. Plasterk, and R.H. Waterston. 1995. The nematode *Caenorhabditis elegans* and its genome *Science* 270: 410-414. (PubMed)

3. Protein Structure and Function

Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. They function as catalysts, they transport and store other molecules such as oxygen, they provide mechanical support and immune protection, they generate movement, they transmit nerve impulses, and they control growth and differentiation. Indeed, much of this text will focus on understanding what proteins do and how they perform these functions.

Several key properties enable proteins to participate in such a wide range of functions.

1. *Proteins are linear polymers built of monomer units called amino acids*. The construction of a vast array of macromolecules from a limited number of monomer building blocks is a recurring theme in biochemistry. Does protein function depend on the linear sequence of amino acids? The function of a protein is directly dependent on its threedimensional structure (Figure 3.1). Remarkably, proteins spontaneously fold up into three-dimensional structures that are determined by the sequence of amino acids in the protein polymer. Thus, *proteins are the embodiment of the transition from the one-dimensional world of sequences to the three-dimensional world of molecules capable of diverse activities*.

2. *Proteins contain a wide range of functional groups*. These functional groups include alcohols, thiols, thioethers, carboxylic acids, carboxamides, and a variety of basic groups. When combined in various sequences, this array of functional groups accounts for the broad spectrum of protein function. For instance, the chemical reactivity associated with these groups is essential to the function of *enzymes*, the proteins that catalyze specific chemical reactions in biological systems (see Chapters 8 – 10).

3. *Proteins can interact with one another and with other biological macromolecules to form complex assemblies*. The proteins within these assemblies can act synergistically to generate capabilities not afforded by the individual component proteins (Figure 3.2). These assemblies include macro-molecular machines that carry out the accurate replication of DNA, the transmission of signals within cells, and many other essential processes.

4. *Some proteins are quite rigid, whereas others display limited flexibility*. Rigid units can function as structural elements in the cytoskeleton (the internal scaffolding within cells) or in connective tissue. Parts of proteins with limited flexibility may act as hinges, springs, and levers that are crucial to protein function, to the assembly of proteins with one another and with other molecules into complex units, and to the transmission of information within and between cells (Figure 3.3).

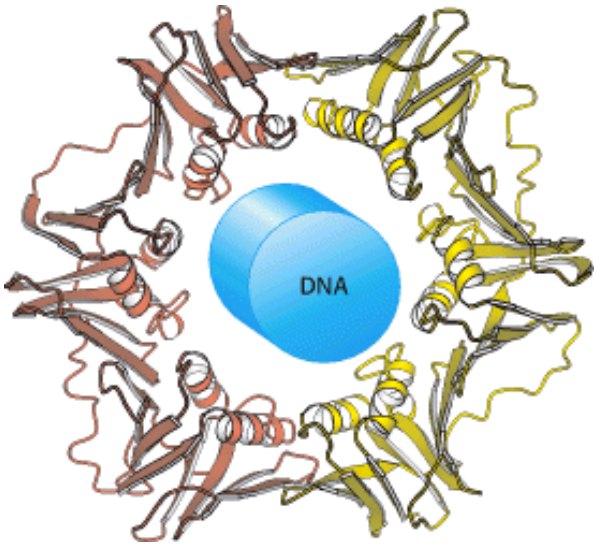


Figure 3.1. Structure Dictates Function. A protein component of the DNA replication machinery surrounds a section of DNA double helix. The structure of the protein allows large segments of DNA to be copied without the replication machinery dissociating from the DNA.

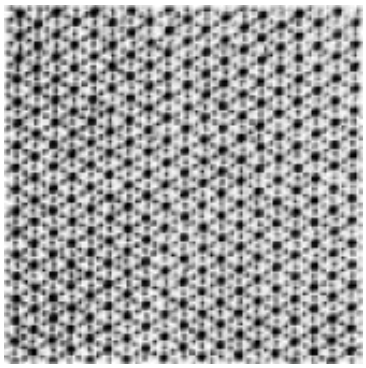


Figure 3.2. A Complex Protein Assembly. An electron micrograph of insect flight tissue in cross section shows a hexagonal array of two kinds of protein filaments. [Courtesy of Dr. Michael Reedy.]

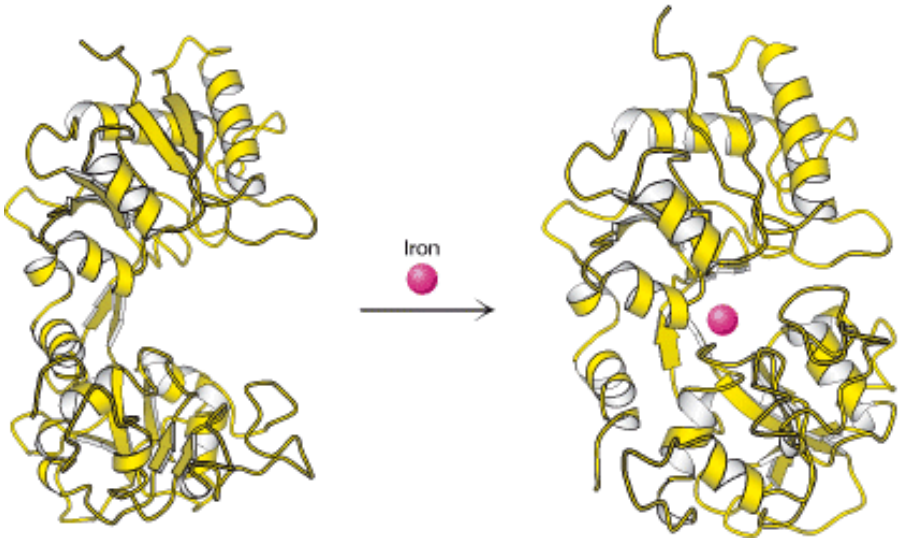
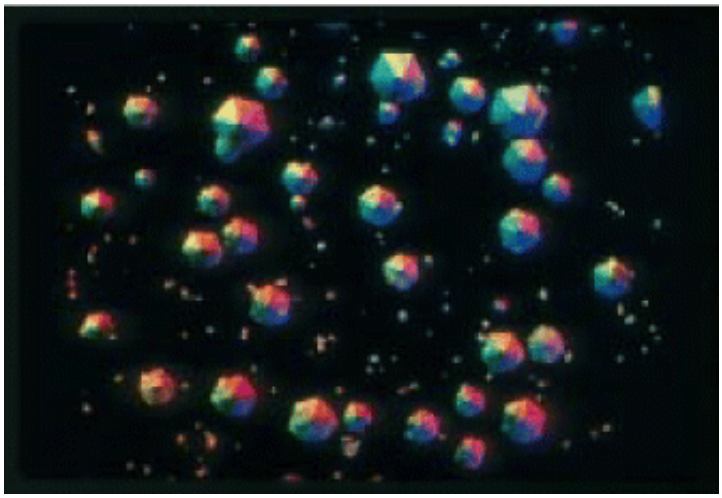
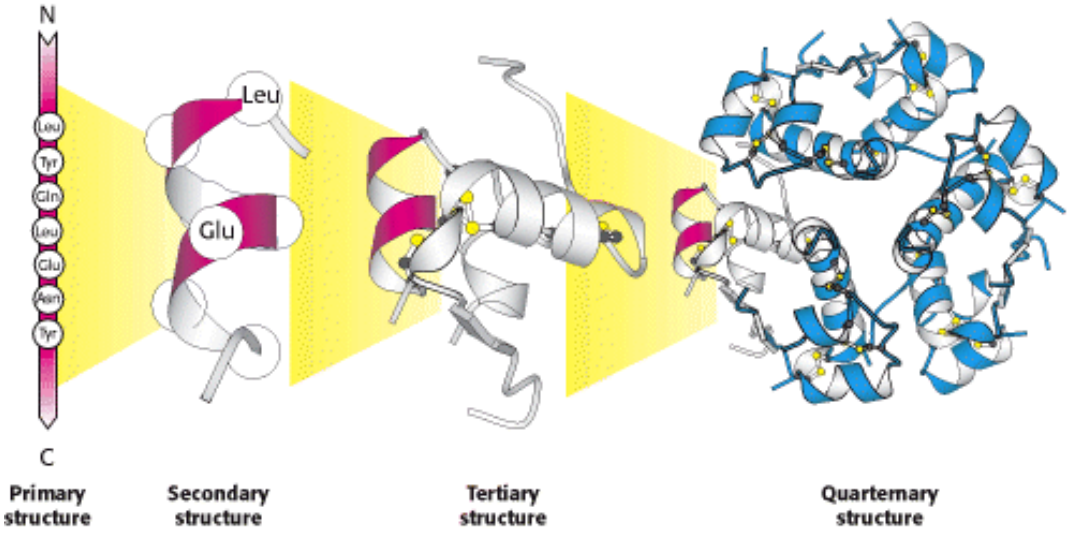


Figure 3.3. Flexibility and Function. Upon binding iron, the protein lactoferrin undergoes conformational changes that

allow other molecules to distinguish between the iron-free and the iron-bound forms.





Crystals of human insulin. Insulin is a protein hormone, crucial for maintaining blood sugar at appropriate levels. (Below) Chains of amino acids in a specific sequence (the primary structure) define a protein like insulin. These chains fold into well-defined structures (the tertiary structure)—in this case a single insulin molecule. Such structures assemble with other chains to form arrays such as the complex of six insulin molecules shown at the far right (the quarternary structure). These arrays can often be induced to form well-defined crystals (photo at left), which allows determination of these structures in detail.[(Left) Alfred Pasieka/Peter Arnold.]

3.1. Proteins Are Built from a Repertoire of 20 Amino Acids

Amino acids are the building blocks of proteins. An α-*amino acid* consists of a central carbon atom, called the α *carbon*, linked to an amino group, a carboxylic acid group, a hydrogen atom, and a distinctive R group. The R group is often referred to as the *side chain.* With four different groups connected to the tetrahedral α-carbon atom, α-amino acids are *chiral;* the two mirror-image forms are called the l isomer and the d isomer (Figure 3.4).

Notation for distinguishing stereoisomers

The four different substituents of an asymmetric carbon atom are assigned a priority according to atomic number. The lowest-priority substituent, often hydrogen, is pointed away from the viewer. The configuration about the carbon is called *S*, from the Latin *sinis-ter* for "left," if the progression from the highest to the lowest priority is counterclockwise. The configuration is called *R*, from the Latin *rectus* for "right," if the progression is clockwise.

Only l amino acids are constituents of proteins. For almost all amino acids, the l isomer has *S* (rather than *R*) absolute configuration (Figure 3.5). Although considerable effort has gone into understanding why amino acids in proteins have this absolute configuration, no satisfactory explanation has been arrived at. It seems plausible that the selection of l over d was arbitrary but, once made, was fixed early in evolutionary history.

Amino acids in solution at neutral pH exist predominantly as *dipolar ions* (also called *zwitterions*). In the dipolar form, the amino group is protonated ($-NH_3^+$) and the carboxyl group is deprotonated ($-COO^-$). The ionization state of an amino acid varies with pH (Figure 3.6). In acid solution (e.g., pH 1), the amino group is protonated ($-NH_3^+$) and the carboxyl group is not dissociated (-COOH). As the pH is raised, the carboxylic acid is the first group to give up a proton, inasmuch as its pK_a is near 2. The dipolar form persists until the pH approaches 9, when the protonated amino group loses a proton. For a review of acid-base concepts and pH, see the appendix to this chapter.

Twenty kinds of side chains varying in *size*, *shape*, *charge*, *hydrogen-bonding capacity*, *hydrophobic character*, and *chemical reactivity* are commonly found in proteins. Indeed, all proteins in all species—bacterial, archaeal, and eukaryotic—are constructed from the same set of 20 amino acids. This fundamental alphabet of proteins is several billion years old. The remarkable range of functions mediated by proteins results from the diversity and versatility of these 20 building blocks. Understanding how this alphabet is used to create the intricate three-dimensional structures that enable proteins to carry out so many biological processes is an exciting area of biochemistry and one that we will return to in Section 3.6.

Let us look at this set of amino acids. The simplest one is *glycine*, which has just a hydrogen atom as its side chain. With two hydrogen atoms bonded to the α-carbon atom, glycine is unique in being *achiral*. *Alanine*, the next simplest amino acid, has a methyl group ($-CH_3$) as its side chain (Figure 3.7).

Larger hydrocarbon side chains are found in *valine*, *leucine*, and *isoleucine* (Figure 3.8). *Methionine* contains a largely *aliphatic* side chain that includes a *thioether* (-S-) group. The side chain of isoleucine includes an additional chiral center; only the isomer shown in Figure 3.8 is found in proteins. The larger aliphatic side chains are *hydrophobic* —that is, they tend to cluster together rather than contact water. The three-dimensional structures of water-soluble proteins are stabilized by this tendency of hydrophobic groups to come together, called *the hydrophobic effect* (see Section 1.3.4). The different sizes and shapes of these hydrocarbon side chains enable them to pack together to form compact structures with few holes. *Proline* also has an aliphatic side chain, but it differs from other members of the set of 20 in that its side chain is bonded to both the nitrogen and the α-carbon atoms (Figure 3.9). Proline markedly influences protein architecture because its ring structure makes it more conformationally restricted than the other amino acids.

Three amino acids with relatively simple *aromatic side chains* are part of the fundamental repertoire (Figure 3.10). *Phenylalanine*, as its name indicates, contains a phenyl ring attached in place of one of the hydrogens of alanine. The aromatic ring of *tyrosine* contains a hydroxyl group. This hydroxyl group is reactive, in contrast with the rather inert side chains of the other amino acids discussed thus far. *Tryptophan* has an indole ring joined to a methylene ($-CH_2-$) group; the indole group comprises two fused rings and an NH group. Phenylalanine is purely hydrophobic, whereas tyrosine and tryptophan are less so because of their hydroxyl and NH groups. The aromatic rings of tryptophan and tyrosine contain

delocalized π electrons that strongly absorb ultraviolet light (Figure 3.11).

A compound's *extinction coefficient* indicates its ability to absorb light. Beer's law gives the absorbance (*A*) of light at a given wavelength:

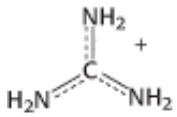
$$A = \epsilon c l \qquad \text{Beer's law}$$

where ε is the extinction coefficient [in units that are the reciprocals of molarity and distance in centimeters (M^{-1} cm^{-1})], *c* is the concentration of the absorbing species (in units of molarity, M), and *l* is the length through which the light passes (in units of centimeters). For tryptophan, absorption is maximum at 280 nm and the extinction coefficient is 3400 M^{-1} cm^{-1} whereas, for tyrosine, absorption is maximum at 276 nm and the extinction coefficient is a less-intense 1400 M^{-1} cm^{-1}. Phenylalanine absorbs light less strongly and at shorter wavelengths. The absorption of light at 280 nm can be used to estimate the concentration of a protein in solution if the number of tryptophan and tyrosine residues in the protein is known.

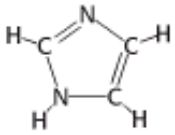
Two amino acids, *serine* and *threonine*, contain aliphatic *hydroxyl groups* (Figure 3.12). Serine can be thought of as a hydroxylated version of alanine, whereas threonine resembles valine with a hydroxyl group in place of one of the valine methyl groups. The hydroxyl groups on serine and threonine make them much more *hydrophilic* (water loving) and *reactive* than alanine and valine. Threonine, like isoleucine, contains an additional asymmetric center; again only one isomer is present in proteins.

Cysteine is structurally similar to serine but contains a *sulfhydryl*, or *thiol* (-SH), group in place of the hydroxyl (-OH) group (Figure 3.13). The sulfhydryl group is much more reactive. Pairs of sulfhydryl groups may come together to form disulfide bonds, which are particularly important in stabilizing some proteins, as will be discussed shortly.

We turn now to amino acids with very polar side chains that render them highly hydrophilic. *Lysine* and *arginine* have relatively long side chains that terminate with groups that are *positively charged* at neutral pH. Lysine is capped by a primary amino group and arginine by a guanidinium group. *Histidine* contains an imidazole group, an aromatic ring that also can be positively charged (Figure 3.14).



Guanidinium



Imidazole

With a pK_a value near 6, the imidazole group can be uncharged or positively charged near neutral pH, depending on its local environment (Figure 3.15). Indeed, histidine is often found in the active sites of enzymes, where the imidazole ring can bind and release protons in the course of enzymatic reactions.

The set of amino acids also contains two with *acidic side chains: aspartic acid* and *glutamic acid* (Figure 3.16). These amino acids are often called *aspartate* and *glutamate* to emphasize that their side chains are usually negatively charged at physiological pH. Nonetheless, in some proteins these side chains do accept protons, and this ability is often functionally important. In addition, the set includes uncharged derivatives of aspartate and glutamate— *asparagine* and *glutamine* —each of which contains a terminal *carboxamide* in place of a carboxylic acid (Figure 3.16).

Seven of the 20 amino acids have readily ionizable side chains. These 7 amino acids are able to donate or accept protons to facilitate reactions as well as to form ionic bonds. Table 3.1 gives equilibria and typical pK_a values for ionization of the side chains of tyrosine, cysteine, arginine, lysine, histidine, and aspartic and glutamic acids in proteins. Two other groups in proteins—the terminal α-amino group and the terminal α- carboxyl group—can be ionized, and typical pK_a

values are also included in Table 3.1.

Amino acids are often designated by either a three-letter abbreviation or a one-letter symbol (Table 3.2). The abbreviations for amino acids are the first three letters of their names, except for asparagine (Asn), glutamine (Gln), isoleucine (Ile), and tryptophan (Trp). The symbols for many amino acids are the first letters of their names (e.g., G for glycine and L for leucine); the other symbols have been agreed on by convention. These abbreviations and symbols are an integral part of the vocabulary of biochemists.

How did this particular set of amino acids become the building blocks of proteins? First, as a set, they are diverse; their structural and chemical properties span a wide range, endowing proteins with the versatility to assume many functional roles. Second, as noted in Section 2.1.1, many of these amino acids were probably available from prebiotic reactions. Finally, excessive intrinsic reactivity may have eliminated other possible amino acids. For example, amino acids such as homoserine and homocysteine tend to form five-membered cyclic forms that limit their use in proteins; the alternative amino acids that are found in proteins—serine and cysteine—do not readily cyclize, because the rings in their cyclic forms are too small (Figure 3.17).

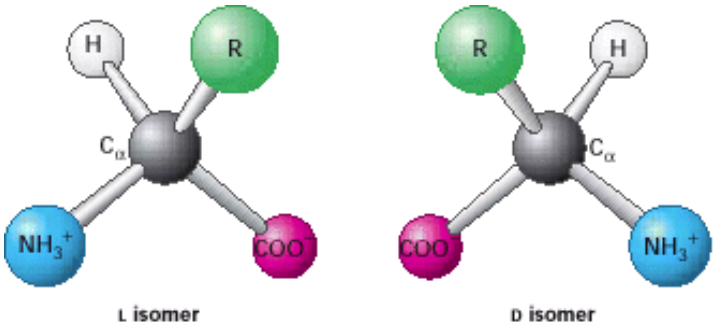


Figure 3.4. The l and d Isomers of Amino Acids. R refers to the side chain. The l and d isomers are mirror images of each other.

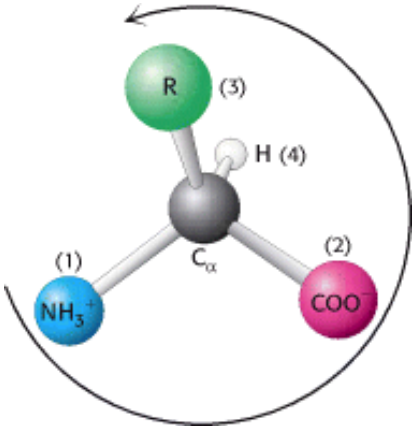


Figure 3.5. Only l Amino Acids Are Found in Proteins. Almost all l amino acids have an *S* absolute configuration (from the Latin *sinister* meaning "left"). The counterclockwise direction of the arrow from highest- to lowest-priority substituents indicates that the chiral center is of the *S* configuration.

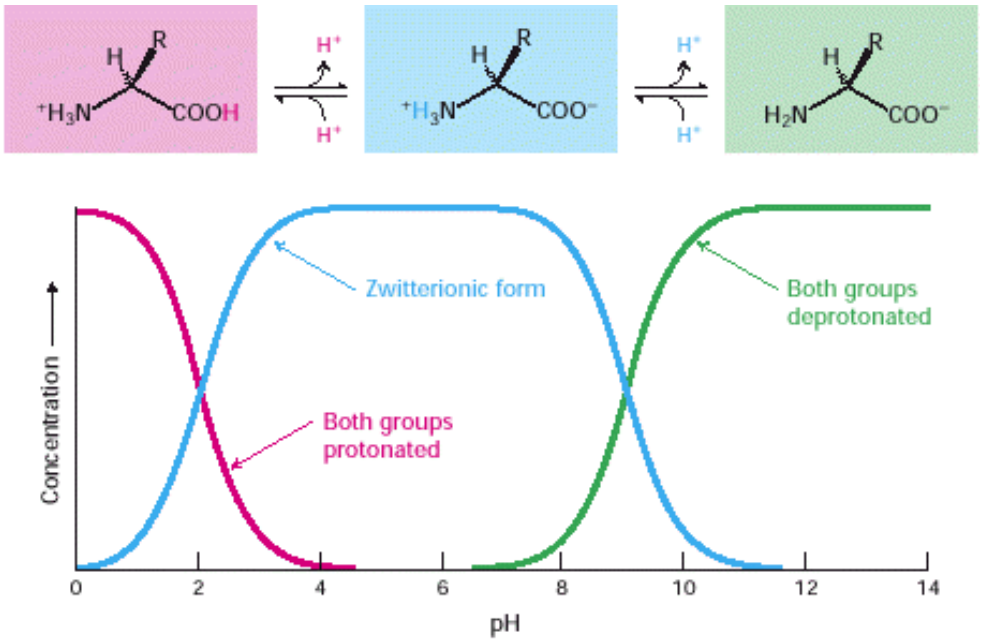


Figure 3.6. Ionization State as a Function of pH. The ionization state of amino acids is altered by a change in pH. The zwitterionic form predominates near physiological pH.

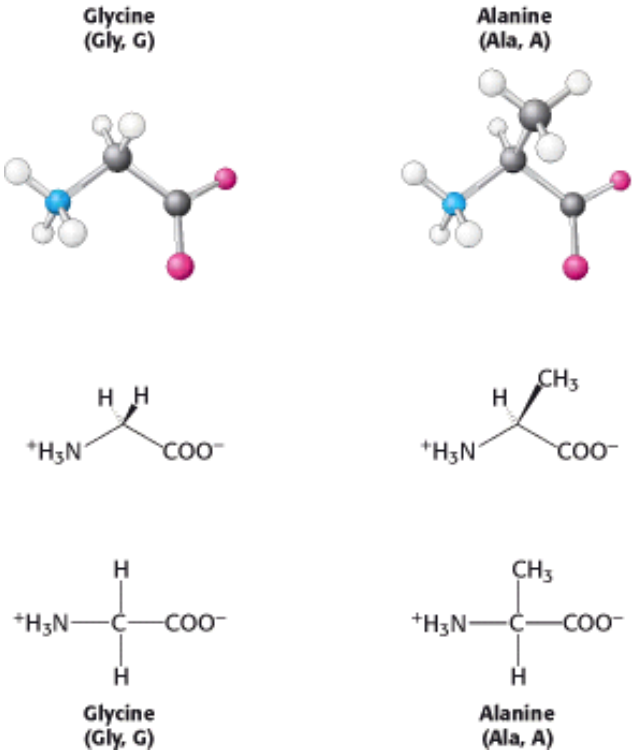


Figure 3.7. Structures of Glycine and Alanine. (Top) Ball-and-stick models show the arrangement of atoms and bonds in space. (Middle) Stereochemically realistic formulas show the geometrical arrangement of bonds around atoms (see Chapters 1 Appendix). (Bottom) Fischer projections show all bonds as being perpendicular for a simplified representation (see Chapters 1 Appendix).

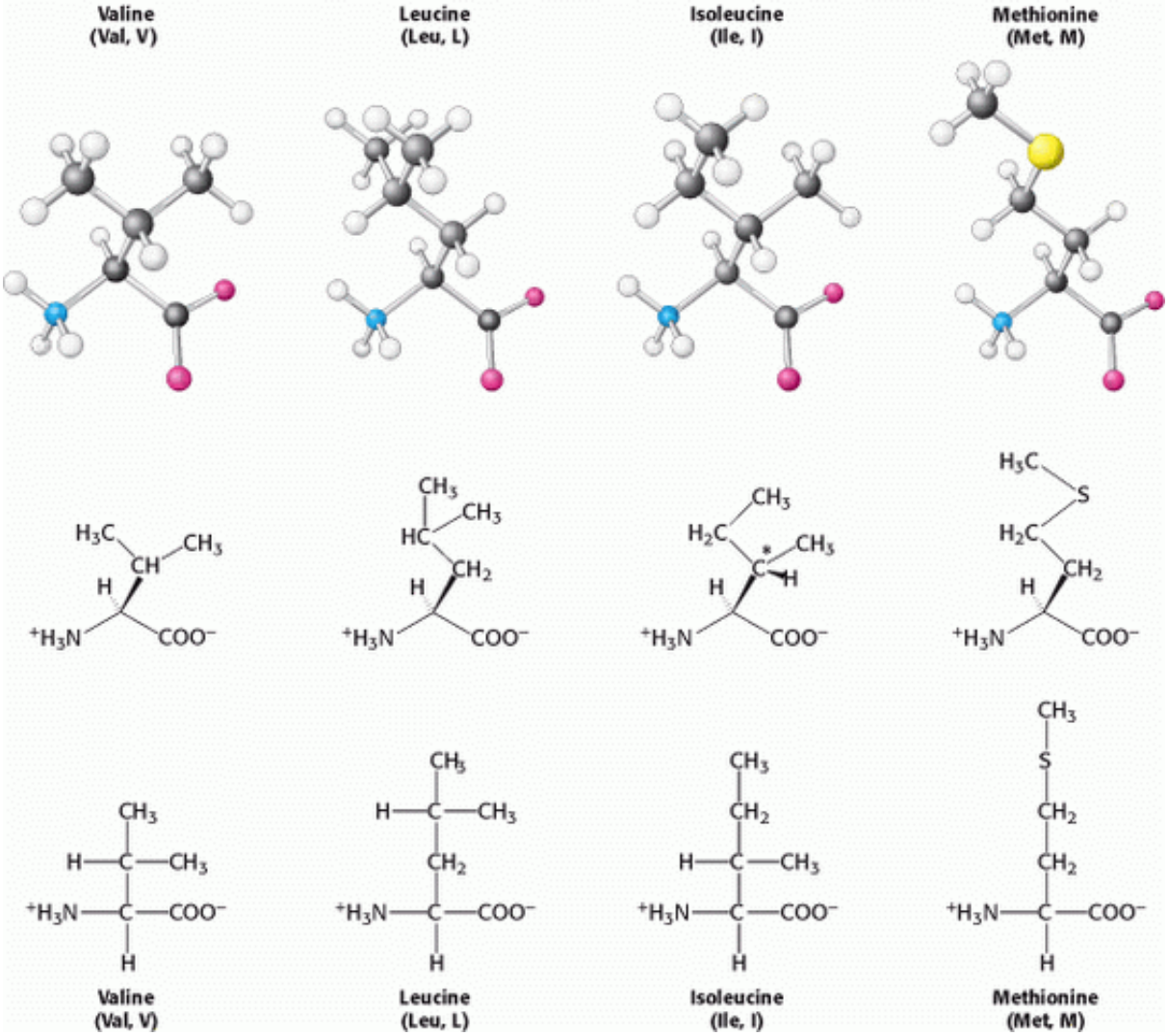


Figure 3.8. Amino Acids with Aliphatic Side Chains. The additional chiral center of isoleucine is indicated by an asterisk.

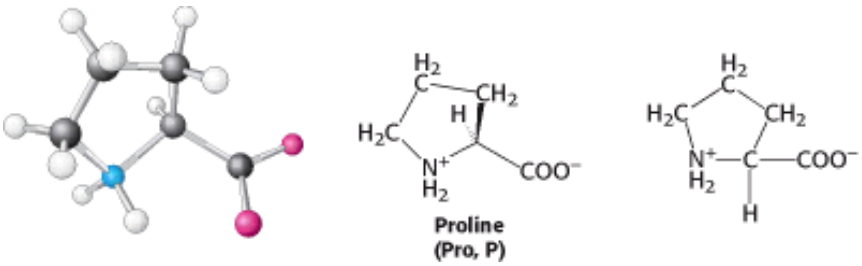


Figure 3.9. Cyclic Structure of Proline. The side chain is joined to both the α carbon and the amino group.

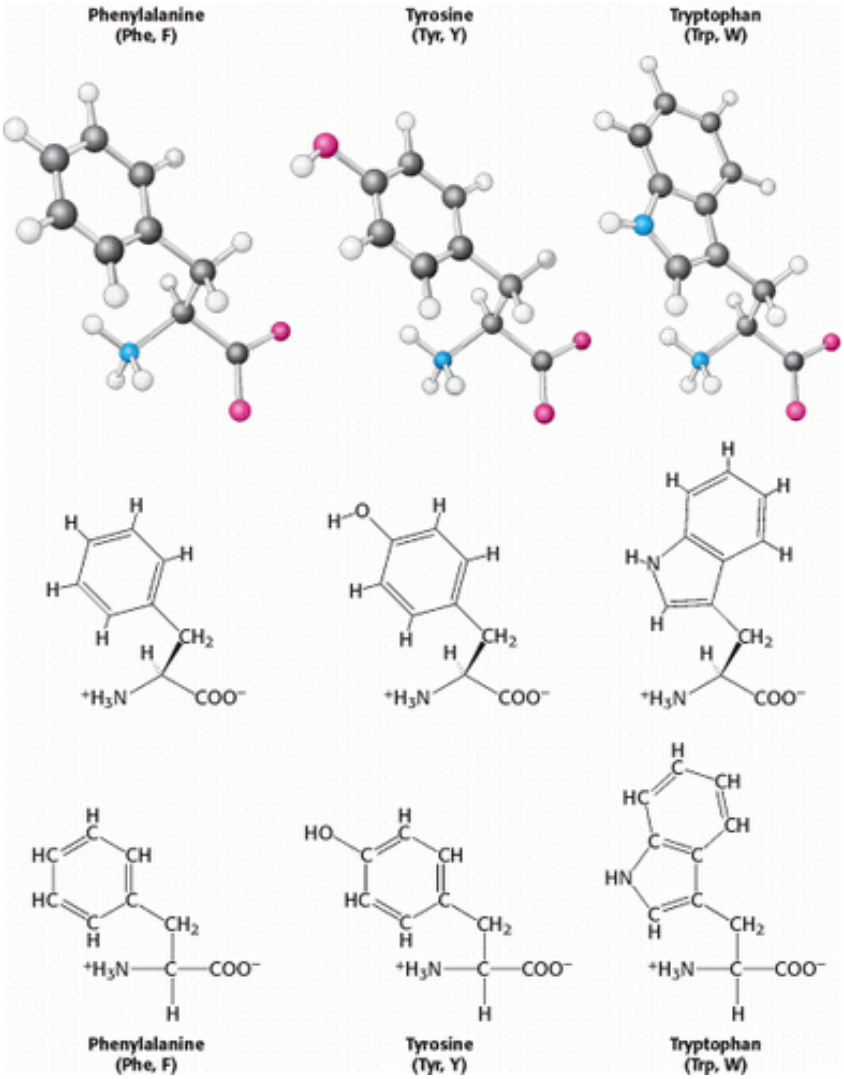


Figure 3.10. Amino Acids with Aromatic Side Chains. Phenylalanine, tyrosine, and tryptophan have hydrophobic character. Tyrosine and tryptophan also have hydrophilic properties because of their -OH and -NH- groups, respectively.

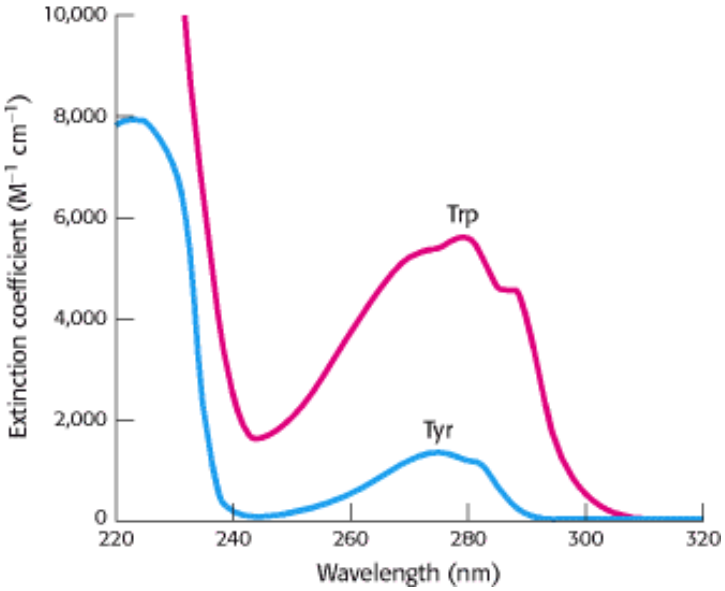


Figure 3.11. Absorption Spectra of the Aromatic Amino Acids Tryptophan (Red) and Tyrosine (Blue). Only these

amino acids absorb strongly near 280 nm. [Courtesy of Greg Gatto].

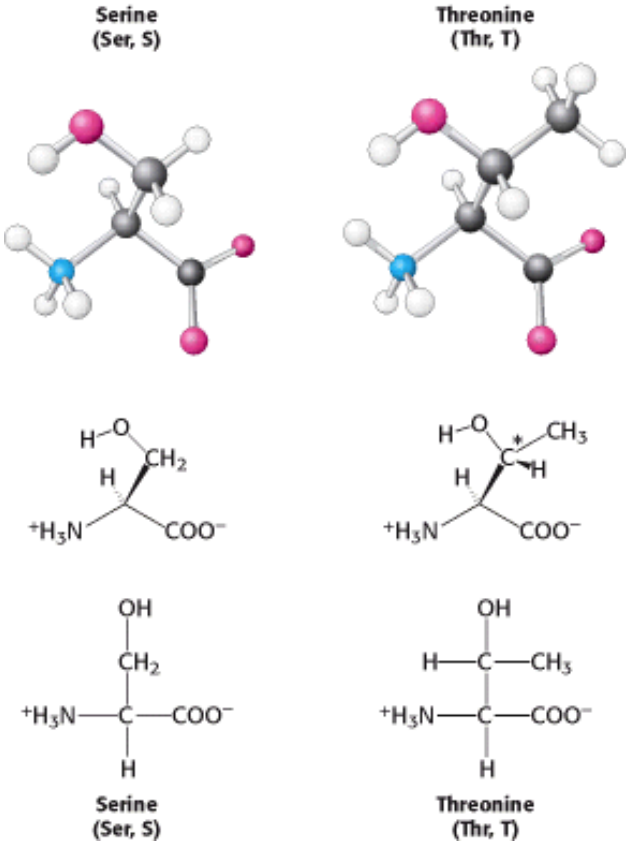


Figure 3.12. Amino Acids Containing Aliphatic Hydroxyl Groups. Serine and threonine contain hydroxyl groups that render them hydrophilic. The additional chiral center in threonine is indicated by an asterisk.

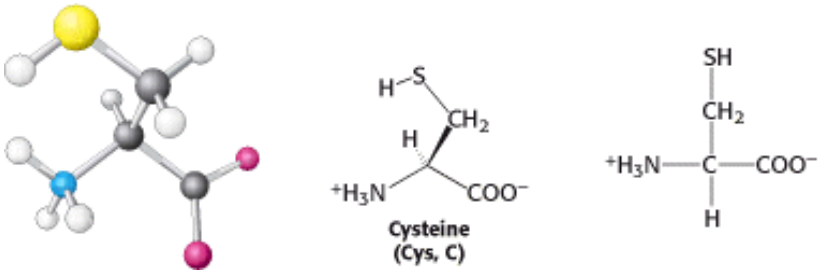


Figure 3.13. Structure of Cysteine.

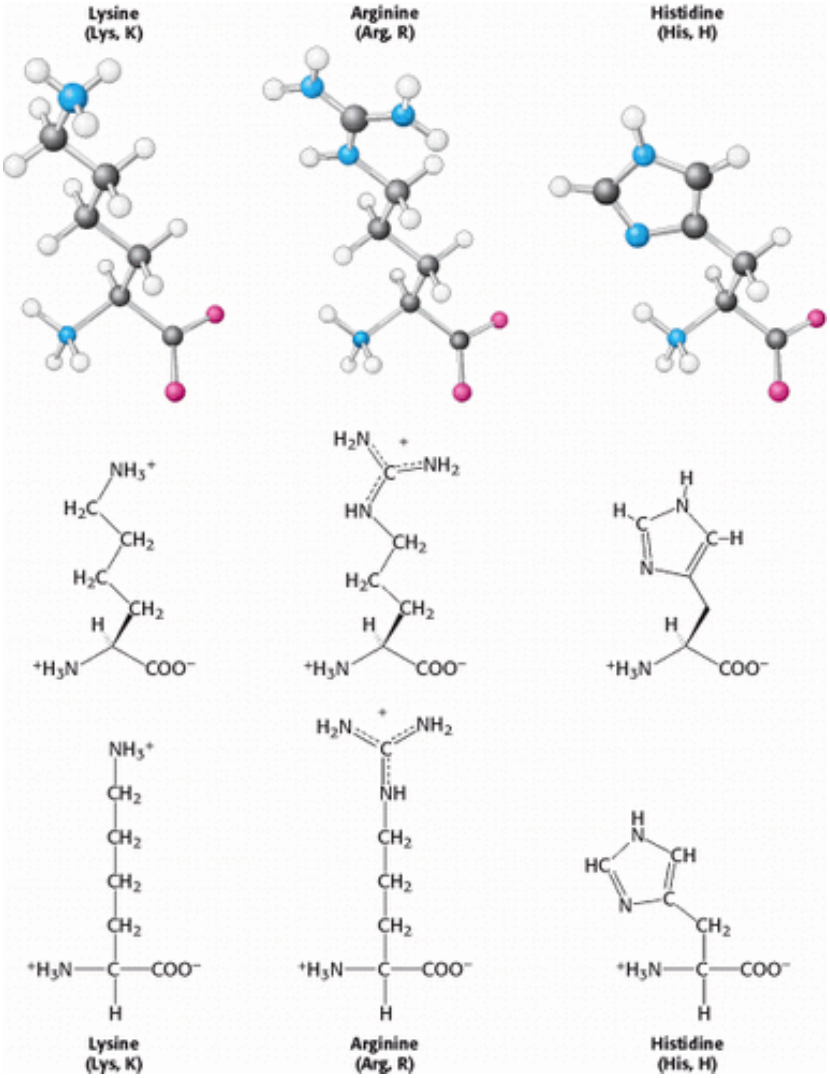


Figure 3.14. The Basic Amino Acids Lysine, Arginine, and Histidine.

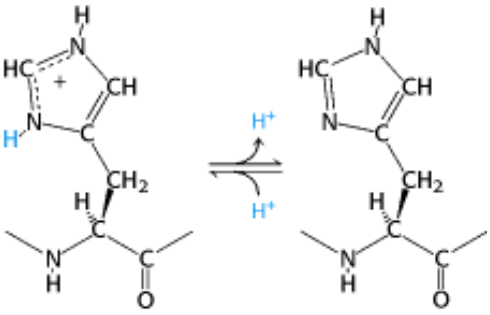


Figure 3.15. Histidine Ionization. Histidine can bind or release protons near physiological pH.

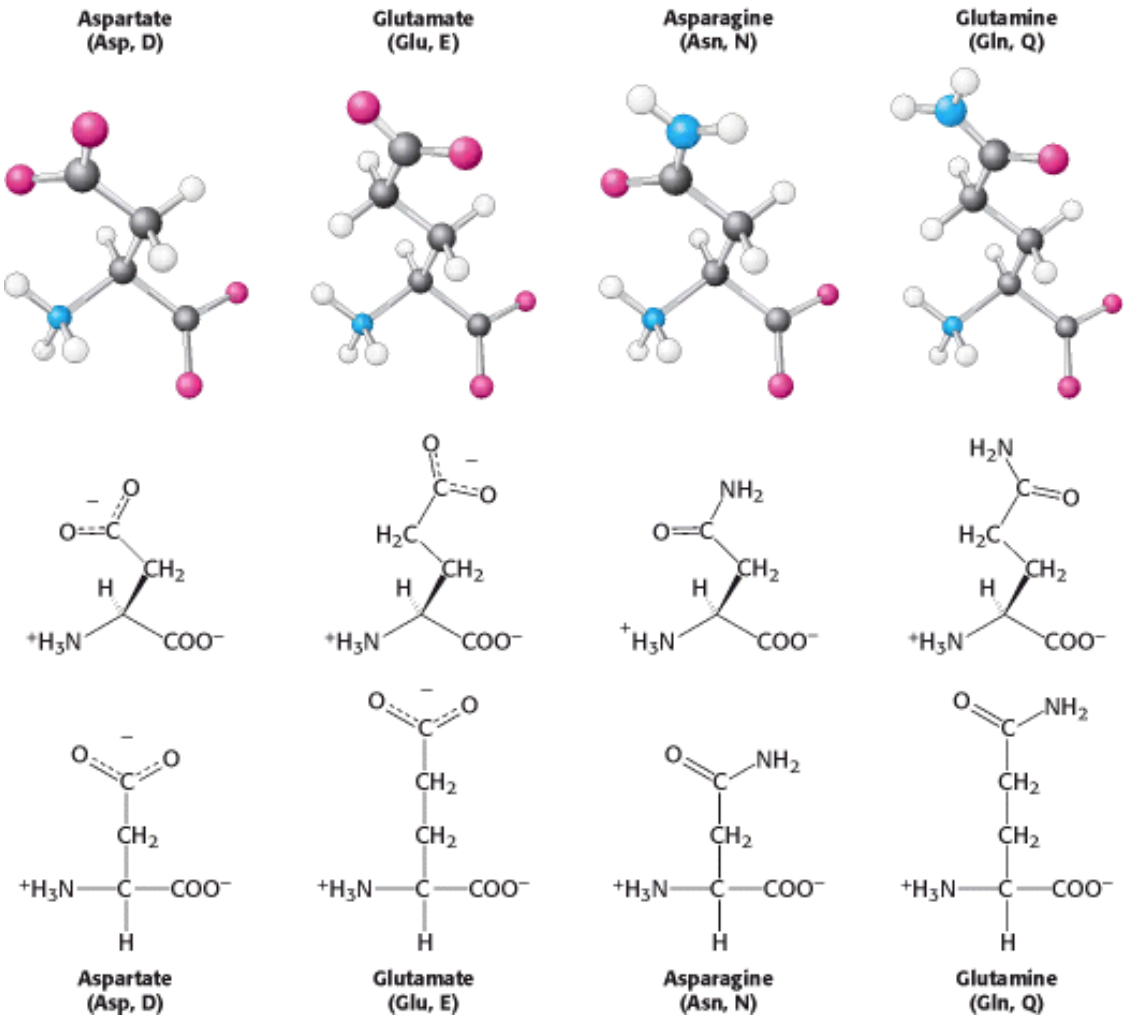


Figure 3.16. Amino Acids with Side-Chain Carboxylates and Carboxamides.

Table 3.1. Typical pK_a values of ionizable groups in proteins