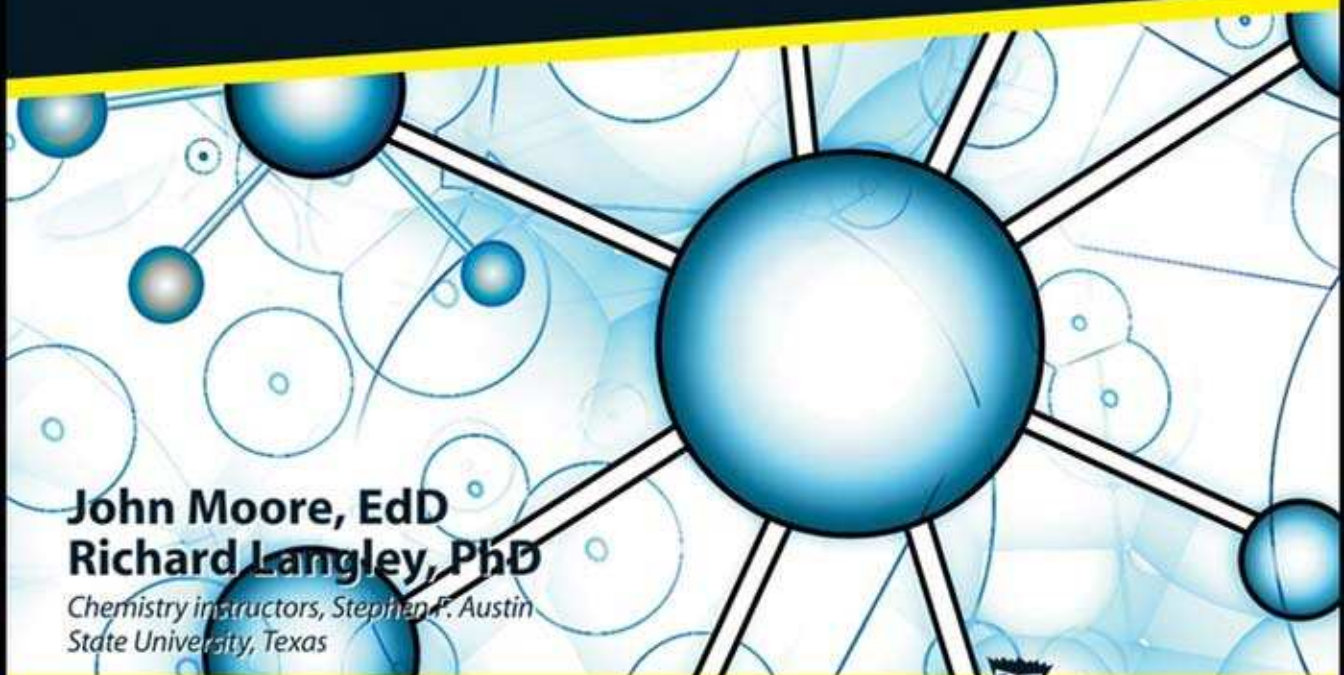
The fun and easy way® to get a handle on biochemistry and raise your grades

Biochemistry

FOR DUMMIES®



John Moore, EdD
Richard Langley, PhD
Chemistry instructors, Stephen F. Austin State University, Texas

A Reference for the Rest of Us!®

FREE eTips at dummies.com®

Biochemistry FOR DUMMIES®

by John T. Moore, EdD and Richard Langley, PhD

Wiley Publishing, Inc.

Biochemistry For Dummies®

Published by
Wiley Publishing, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Published by Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

Library of Congress Control Number: 2008923128

ISBN: 978-0-470-19428-7

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

About the Authors

John Moore grew up in the foothills of Western North Carolina. He attended the University of North Carolina at Asheville, where he received his bachelor's degree in chemistry. He earned his master's degree in chemistry from Furman University in Greenville, South Carolina. After a stint in the United States Army, he decided to try his hand at teaching. In 1971, he joined the chemistry faculty of Stephen F. Austin State University in Nacogdoches, Texas, where he still teaches chemistry. In 1985 he started back to school part time and in 1991 received his doctorate in education from Texas A&M University. For the last five years has been the co-editor (along with one of his former students) of the "Chemistry for Kids" feature of *The Journal of Chemical Education*. In 2003, his first book, *Chemistry For Dummies*, was published, soon to be followed by *Chemistry Made Simple*. John enjoys cooking and making custom knife handles from exotic woods.

Richard Langley grew up in southwestern Ohio. He attended Miami University in Oxford, Ohio, where he received bachelor's degrees in chemistry and mineralogy and then a master's degree in chemistry. His next stop was the University of Nebraska, where he received his doctorate in chemistry. Afterwards, he took a postdoctoral position at Arizona State University in Tempe, Arizona, followed by a visiting assistant professor position at the University of Wisconsin at River Falls. In 1982, he moved to Stephen F. Austin State University. For the past several years, he and John have been graders for the Free Response portion of the AP Chemistry Exam. He and John have collaborated on several writing projects, including *5 Steps To A 5 on the AP: Chemistry* and *Chemistry for the Utterly Confused*. Rich enjoys jewelry making and science fiction.

Dedication

To my wife, Robin; sons, Matthew and Jason; my wonderful daughter-in-law, Sara; and the two most wonderful grandkids in the world, Zane and Sadie. I love you guys. — John

To my mother. — Rich

Authors' Acknowledgments

We would not have had the opportunity to write this book without the encouragement of our agent, Grace Freedson. We would also like to thank Kristin DeMint for her support and assistance in the early portion of this project and to Corbin Collins who helped us complete it. Thanks to our colleague Michele Harris who helped with suggestions and ideas. And many thanks to all of the people at Wiley publishing who helped bring this project from concept to publication.

Publisher's Acknowledgments

We're proud of this book; please send us your comments through our Dummies online registration form located at www.dummies.com/register/.

Some of the people who helped bring this book to market include the following:

Acquisitions, Editorial, and Media Development

Project Editors: Kristin DeMint and Corbin Collins

Acquisitions Editor: Stacy Kennedy

Copy Editors: Josh Dials and Corbin Collins

Technical Editor: Medhane Cumbay

Project Manager I: Laura Moss-Hollister

Media Development Specialist: Kit Malone

Editorial Manager: Michelle Hacker

Editorial Supervisor and Reprint Editor: Carmen Krikorian

Art Coordinator: Alicia B. South

Editorial Assistants: David Lutton and Leeann Harney

Cover Photos: Rox Woodward

Cartoons: Rich Tennant (www.the5thwave.com)

Composition Services

Project Coordinator: Lynsey Stanford

Layout and Graphics: Carrie A. Cesavice, Alissa D. Ellet, Stephanie D. Jumper

Proofreaders: Laura Albert, Caitie Kelly, Toni Settle

Indexer: Potomac Indexing, LLC

Publishing and Editorial for Consumer Dummies

Diane Graves Steele, Vice President and Publisher, Consumer Dummies

Joyce Pepple, Acquisitions Director, Consumer Dummies

Kristin A. Cocks, Product Development Director, Consumer Dummies

Kathleen Nebenhaus, Vice President and Executive Publisher, Consumer Dummies, Lifestyles, Pets, Education Publishing for Technology Dummies

Composition Services

Gerry Fahey, Vice President of Production Services

Debbie Stailey, Director of Composition Services

Contents at a Glance

Table of Contents

Introduction

Welcome to *Biochemistry For Dummies!*

We are certainly happy you have decided to delve into the fascinating world of biochemistry. Biochemistry is a complex area of chemistry, but understanding biochemistry isn't really complex. It takes hard work, attention to detail, and the desire to know and to imagine. Biochemistry, like any area of chemistry, is not a spectator sport. You must interact with the material, try different explanations, and ask yourself why things happen the way they do.

Work hard and you will get through your biochem course. More importantly, you might grow to appreciate the symphony of chemical reactions that take place within a living organism, whether it be a one-celled organism, a tree, or a person. As each individual instrument contributes to an orchestra, each chemical reaction is necessary, and sometimes its part is quite complex. However, when you combine all the instruments, and each instrument functions well, the result can be a wonder to behold and hear. If one or two instruments are a little out of tune or aren't being played well, the orchestra still functions — but things are a little off. The sound isn't quite as beautiful, or there is a nagging sensation of something being wrong. The same is true of an organism. If all the reactions occur correctly at the right time, the organism functions well. If a reaction or a few reactions are off in some way, the organism may not function nearly as well. Genetic diseases, electrolyte imbalance, and other problems may cause the organism to falter. And what happens then? Biochemistry is often where ways of restoring the organism to health are found.

About This Book

Biochemistry For Dummies is an overview of the material covered in a typical college-level biochemistry course. We have made every attempt to keep the material as current as possible, but the field is changing ever so quickly. The basics, however, stay the same, and that is where we concentrate our efforts. We also include information on some of the applications of biochemistry that you read about in your everyday life, such as forensics, cloning, gene therapy, genetic testing, genetically modified foods, and so on.

As you flip through this book you will see a lot of chemical structures and reactions. Much of the biochemistry revolves around knowing the structures of the molecules involved in biochemical reactions. If you are in a biochemistry course, you probably have had at least one semester of organic chemistry. You will recognize many of the structures, or at least the functional groups, from your study of organic chem. You will see many of those mechanisms that you loved/hated here in biochemistry.

If you bought this book just to gain general knowledge about a fascinating subject, try not to get bogged down in the details. Skim the chapters. If you find a topic that interests you, stop and dive in. Have fun learning something new.

Conventions Used in This Book

We have organized this text in a logical progression of topics that might be used in a biochemistry course. We have made extensive use of structures and reactions. While reading, try to follow along in the associated figures, whether they be structures or reactions. The icons point out things to which you should pay particular attention, for various reasons. If you are taking a biochemistry course, use this rather inexpensive book to supplement that very expensive biochemistry textbook.

Icons Used in This Book

If you have ever read other *For Dummies* books (such as the wonderful *Chemistry For Dummies*) you will recognize the icons used in this book, but here are their meanings anyway:

This icon is a flag for those really important things that you shouldn't forget as you go deeper into the world of biochemistry.

We use this icon to alert you to a tip on the easiest or quickest way to learn a concept. Between the two of us, we have almost 70 years of teaching experience. We've learned a few tricks along the way and we don't mind sharing.

The Real World icon points out information that has direct application in the everyday world.

The Warning icon points to a procedure or potential outcome that can be dangerous. We call it our Don't-Try-This-At-Home icon.

What You're Not to Read

Don't read what you don't need. Concentrate on the area(s) in which you need help. If you are interested in real-world applications of biochemistry, by all means read those sections (indicated by the Real World icon). However, if you just need help on the straight biochemistry, feel free to skip the applications. You don't have a whole lot of money invested in this book, so don't feel obligated to read everything. When you're done, you can put it in your bookshelf alongside *Chemistry For Dummies, The Doctor Who Error Finder,* and *A Brief History of Time* as a conversation piece.

Foolish Assumptions

We assume — and we all know about the perils of assumptions — that you are one of the following:

- A student taking a college-level biochemistry course.
- A student reviewing your biochemistry for some type of standardized exam (the MCAT, for example).
- An individual who just wants to know something about biochemistry.
- A person who has been watching way too many forensic TV shows.

If you fall into a different category, we hope you enjoy this book anyway.

How This Book Is Organized

Here is a very brief overview of the topics we cover in the various parts of this book. Use these descriptions and the Table of Contents to map out your strategy of study.

Part I: Setting the Stage: Basic Biochemistry Concepts

This part deals with basic aspects of chemistry and biochemistry. In the first chapter you find out about the field of biochemistry and its relationship to other fields within chemistry and biology. You also get a lot of info about the different types of cells and their parts. In Chapter 2 we review some aspects of water chemistry that have direct applications to the field of biochemistry, including pH and buffers. Finally, you end up with a one-chapter review of organic chemistry, from functional groups to isomers.

Part II: The Meat of Biochemistry: Proteins

In this part we concentrate on proteins. You are introduced to amino acids, the building blocks of proteins. Having the building blocks in hand, in the next chapter we show you the basics of amino acid sequencing and the different types of protein structure. Finally, we will finish this part with a discussion of enzyme kinetics, both catalysts (speeding up reactions) and inhibitors (slowing them down).

Part III: Carbohydrates, Lipids, Nucleic Acids, and More

In this part we show you a number of biochemical species. You'll see that carbohydrates are far more complex than that doughnut you just ate might lead you to believe, but we do show you some biochemistry that is just sweet! Then we jump over to lipids and steroids. Next are nucleic acids and the genetic code (Da Vinci, eat your heart out!) of life with DNA and RNA. Then it's on to vitamins (they are involved more than once a day) and hormones (no humor here — it would be just too easy).

Part IV: Bioenergetics and Pathways

It all comes down to energy, one way or another. In these chapters we look at energy requirement and where that energy goes. This is where you meet our friend ATP and battle the formidable Citric Acid Cycle. Finally, since you will be hot and sweaty anyway, we throw you into the really smelly bog of nitrogen chemistry.

Part V: Genetics: Why We Are What We Are

In this part we tell you all about making more DNA, the processes of replication, and several of the applications related to DNA sequencing. Then it's off to RNA and protein synthesis. We also spend some time talking about the Human Genome Project.

Part VI: The Part of Tens

In this final part of the book we discuss ten great applications of biochemistry to the everyday world and reveal ten not-so-typical biochemical careers.

Where to Go from Here

The answer to this question really depends of your prior knowledge and goals. As with all *For Dummies* books, this one attempts to make all the chapters independent, so that you can pick a chapter containing material you are having difficulty with and get after it, without having to have read other chapters first. If you feel comfortable with the topics covered in general and organic chemistry, feel free to skip Part I. If you want a general overview of biochemistry, skim the remainder of the book. Dive deeper into the gene pool when you find a topic that interests you.

And for all of you, no matter who you are or why you are reading this book, we hope that you have fun reading it and that it helps you to learn biochemistry.

Part I

Setting the Stage: Basic Biochemistry Concepts

"Look – I'm only going to do this once, but it should help you remember the nonpolar hydrophobic principle and its effect on the surface tension of water."

In this part . . .

We go over some basic aspects of chemistry, organic chemistry, and biochemistry. First we survey the field of biochemistry and its relationship to other disciplines within chemistry and biology. We cover several different types of cells and their parts. Then we look at some features of water chemistry that apply to biochemistry, paying attention to pH and buffers. In the end, you get a brush-up on your organic chemistry, which sets the stage for Part II.

Chapter 1

Biochemistry: What You Need to Know and Why

In This Chapter

- Considering biochemistry
- Finding out about the types of cells
- Seeing the differences between plant and animal cells

If you are enrolled in a biochemistry course, you may want to skip this chapter and go right to the chapter(s) where we discuss the material you are having trouble with. But if you are *thinking* about taking a course in biochemistry or just want to explore an area that you know little about, keep reading. This chapter gives you basic information about cell types and the parts of the cell — which are extremely important in biochemistry.

Sometimes it's easy to get lost in the technical stuff and forget about the big picture. This chapter sets the stage for the details.

Why Biochemistry?

We suppose the flippant answer would be "Why not?" or "Because it is required."

That first response is not too bad an answer, actually. Look around. See all the living or once living things around you? The processes that allow them to grow, multiply, age, and die are all biochemical in nature. Sometimes we sit back and marvel at the complexity of life, the myriad of chemical reactions that are taking place right now within our own bodies, how all these biochemical reactions are working together so that we can sit and contemplate them. When John learned about the minor structural difference between starch and cellulose he remembers thinking: "Just that little difference in the one linkage

between those units is basically the difference between a potato and a tree?" It made him want to learn more, to delve into the complexity of the chemistry of living things, to try to understand. We encourage you to step back from the details occasionally and marvel at the complexity and beauty of life.

What Is Biochemistry and Where Does It Take Place?

Biochemistry is the chemistry of living organisms. Biochemists study the chemical reactions that occur at the molecular level of organisms. Normally it is listed as a separate field of chemistry. However, in some schools it is part of biology, and in others it is separate from both chemistry and biology.

Biochemistry really reaches out and combines aspects of all the fields of chemistry. Because carbon is the element of life, *organic chemistry* plays a large part in biochemistry. Many times biochemists study how fast reactions occur — that's *physical chemistry.* Often metals are incorporated into biochemical structures (such as iron in hemoglobin) — that's *inorganic chemistry.* Biochemists use sophisticated instrumentation to determine amounts and structures — that's *analytical chemistry.* Biochemistry is similar to *molecular biology;* both study living systems at the molecular level, but biochemists concentrate on the chemical reactions that are occurring.

Biochemists may study individual electron transport within the cell, or they may study the processes involved in digestion. If it's alive, biochemists will study it.

Types of Living Cells

All living organisms contain cells. A *cell* is a prison of sorts. The working apparatus of the cell is imprisoned within the "bars" — known as the *cell membrane*. Just as a prison inmate can still communicate with the outside world, so can the cell contents. The prisoner must be fed, so nutrients must be able to enter every living cell. There is a sanitary system for the elimination of waste. And, just as inmates may work to provide materials for society outside the prison, a cell may produce materials for life outside the cell.

There are two types of cells: prokaryotes and eukaryotes. (Viruses also bear some similarities to cells, but these are limited.) Prokaryotic cells are the simplest type of cells. Many one-celled organisms are prokaryotes.

The simplest way to distinguish these two types is that a *prokaryotic cell* contains no well-defined nucleus, whereas the opposite is true for a *eukaryotic cell.*

Prokaryotes

Prokaryotes are mostly bacteria. Besides the lack of a nucleus, there are few well-defined structures inside a prokaryotic cell. The prison wall has three components: a cell wall, an outer membrane, and a plasma membrane. This wall allows a controlled passage of material into or out of the cell. The materials necessary for proper functioning of the cell float about inside it, in a soup known as the *cytoplasm.* Figure 1-1 depicts a simplified version of a prokaryotic cell.

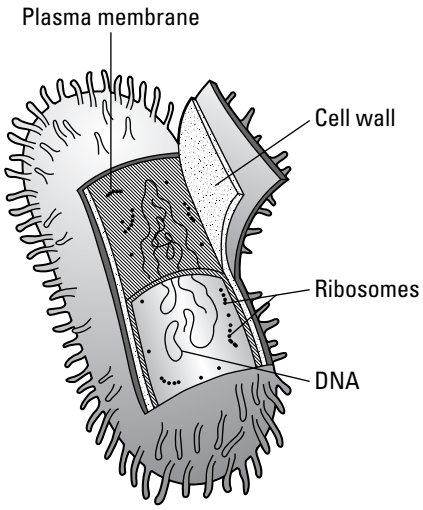


Figure 1-1: Simplified prokaryotic cell.

Eukaryotes

Eukaryotes are animals, plants, fungi, and protists. *You* are a eukaryote. In addition to having a nucleus, eukaryotic cells have a number of membrane-enclosed components known as *organelles*. Eukaryotic organisms may be either unicellular or multicellular. In general, eukaryotic cells contain much more genetic material than prokaryotic cells.

Animal Cells and How They Work

All animal cells (which are, as you now know, eukaryotic cells) have a number of components, most of which are considered to be organelles. The primary components of animal cells are listed in Table 1-1. (These components, and a few others, are also present in plant cells.) Figure 1-2 illustrates a simplified animal cell.

Table 1-1 Parts of an Animal Cell

Cell membrane	Centrioles
Endoplasmic reticulum	Golgi apparatus
Lysosomes	Mitochondria
Nucleus and nucleolus	Ribosomes
Small vacuoles	

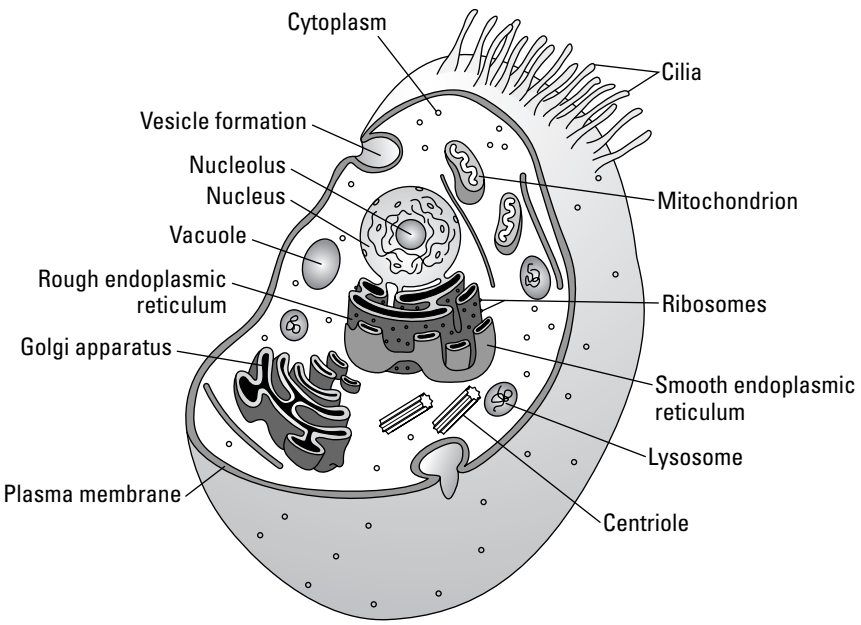


Figure 1-2: Simplified illustration of an animal cell.

The plasma membrane separates the material inside the cell from everything outside the cell. The *plasma* or cytoplasm is the fluid inside the cell. It is important for the health of the cell to prevent this fluid from leaking out. However, necessary materials must be able to enter through the membrane, and other materials, including waste, must be able to exit through the membrane.

Transport through the membrane may be active or passive. *Active transport* requires that a price be paid for a ticket to enter (or leave) the cell. The cost of the ticket is energy. *Passive transport* does not require a ticket. Passive transport methods include diffusion, osmosis, and filtration.

Centrioles behave as the "train conductors" of the cell. They organize microtubles, which help move the parts of the cell during cell division.

The cell can be thought of as a smoothly running factory. The *endoplasmic reticulum* is the main part of the cell factory. There are two basic regions to this structure, known as the *rough* endoplasmic reticulum and the *smooth* endoplasmic reticulum. The rough endoplasmic reticulum contains ribosomes, and the smooth endoplasmic reticulum contains no ribosomes (more about ribosomes and their function is coming up in this chapter). The rough endoplasmic reticulum, through the ribosomes, is the assembly line of the factory. The smooth endoplasmic reticulum is more like the shipping department, which ships the products of the reactions that occur within the cell, to the Golgi apparatus.

The *Golgi apparatus* serves as the postal system of the cell. It looks a bit like a maze, and within it, materials produced by the cell are packaged in vesicles, small membrane-enclosed sacs. The vesicles are then mailed to other organelles or to the cell membrane for export. The cell membrane contains "customs officers" (called *channels*), who allow secretion of the contents from the cell. Secreted substances are then available for other cells or organs.

Lysosomes are the landfills of the cell. They contain digestive enzymes that break down substances that may harm the cell (Chapter 6 has a lot more about enzymes). The products of this digestion may then safely reenter the cell. Lysosomes also digest "dead" organelles. This slightly disturbing process, called *autodigestion,* is really part of the cell digesting itself.

The *mitochondria* (singular mitochondrion) are the cell's power plants, where the cell produces energy. Mitochondria use food, primarily the carbohydrate *glucose,* to produce energy, which comes mainly in the form of *adenosine triphosphate* (ATP — to which Chapter 13 is dedicated).

Each cell has a *nucleus* and, inside it, a *nucleolus.* These serve as the control center of the cell and are the root from which all future generations originate. A double layer known as the *nuclear membrane* surrounds the nucleus. Usually the nucleus contains a mass of material called *chromatin.* If the cell is entering a stage leading to reproducing itself through cell division, the chromatin separates into *chromosomes.*

In addition to conveying genetic information to future generations, the nucleus produces two important molecules for the interpretation of this information. These molecules are *messenger ribonucleic acid* (mRNA) and *transfer ribonucleic acid* (tRNA). The nucleolus produces a third type of ribonucleic acid known as *ribosomal ribonucleic acid* (rRNA). (Chapter 9 is all about nucleic acids.)

Ribosomes contain protein and ribonucleic acid subunits. It is in the ribosomes where the amino acids are assembled into *proteins.* Many of these proteins are enzymes, which are part of nearly every process occurring in the organism. (Part II of this book is devoted to amino acids, proteins, and enzymes.)

The *small vacuoles,* or simply *vacuoles,* serve a variety of functions, including storage and transport of materials. The stored materials may be for later use or may be waste material no longer needed by the cell.

A Brief Look at Plant Cells

Plant cells contain the same components as animal cells — plus a cell wall, a large vacuole, and, in the case of green plants, chloroplasts. Figure 1-3 illustrates a typical plant cell.

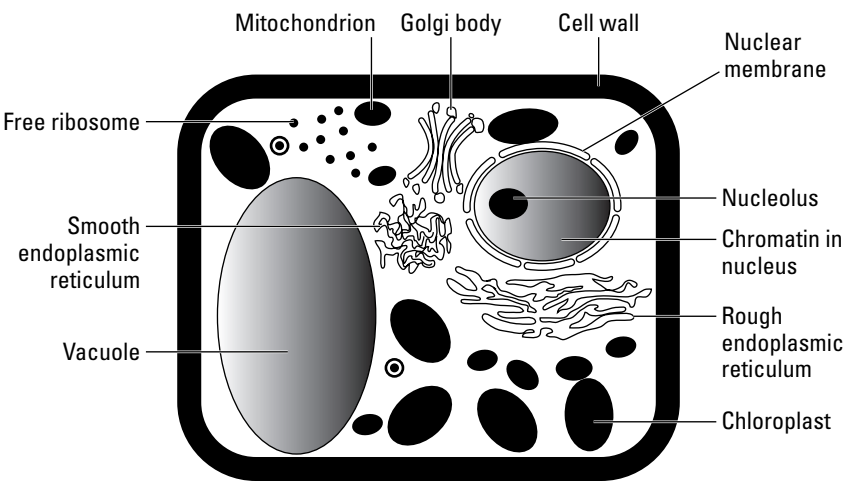


Figure 1-3: Simplified illustration of a plant cell.

The *cell wall* is composed of cellulose. Cellulose, like starch, is a polymer of glucose. The cell wall provides structure and rigidity.

The *large vacuole* serves as a warehouse for large starch molecules. Glucose, which is produced by photosynthesis, is converted to *starch,* a polymer of glucose. At some later time, this starch is available as an energy source. (Chapter 7 talks a lot more about glucose and other carbohydrates.)

Chloroplasts, present in green plants, are specialized chemical factories. These are the sites of photosynthesis, in which *chlorophyll* absorbs sunlight and uses this energy to combine carbon dioxide and water to produce glucose and release oxygen gas.

The green color of many plant leaves is due to the magnesium-containing compound chlorophyll.

Now that you know a little about cells, press on and let's do some biochemistry!

Chapter 2

Dive In: Water Chemistry

In This Chapter

- Understanding the roles and properties of water
- Exploring the differences between acids and bases
- Examining acid-base equilibria with the Brønsted-Lowry theory
- Controlling pH with buffers

Water is one of the most important substances on earth. We swim, bathe, boat, and fish in it. It carries our waste from our homes and is used in the generation of electrical power. We drink it in a variety of forms: pure water, soft drinks, tea, coffee, margaritas, and so on. Water, in one form or another, moderates the temperature of the earth and of our bodies.

In the area of biochemistry, water is also one of the lead actors. Our bodies are about 70 percent water. Water plays a role in the transport of material to and from cells. And many, many aqueous solutions take part in the biochemical reactions in the body.

In this chapter, we examine the structure and properties of the water molecule. We explain how water behaves as a solvent. We look at the properties of acids and bases and the equilibria that they may undergo. Finally, we discuss the pH scale and buffers, including the infamous Henderson-Hasselbalch equation. Sit back, grab a glass of water, and dive in!

The Fundamentals of H_2O

Water is essential to life; in fact, human beings are essentially big sacks of water. Water accounts for 60–95 percent of our living cells, and 55 percent of the water in the human body is in intracellular fluids. The remaining 45 percent (extracellular) is divided between the following:

- Plasma (8 percent)
- Interstitial and lymph (22 percent)
- Connective tissue, cartilage, and bone (15 percent)

Water also is necessary as a solvent for the multitude of biochemical reactions that occur in the body:

- Water acts as a transport medium across membranes, carrying substances into and out of cells.
- Water helps maintain the temperature of the body.
- Water acts as a solvent (carrying dissolved chemicals) in the digestive and waste excretion systems.

Healthy humans have an intake/loss of about two liters of water per day. The intake is about 45 percent from liquids and 40 percent from food, with the remainder coming from the oxidation of food. The loss is about 50 percent from urine and 5 percent from feces, with the remainder leaving through evaporation from the skin and lungs. A water balance must be maintained within the body. If the loss of water significantly exceeds the intake, the body experiences dehydration. If the water loss is significantly less than the intake, water builds up in the body and causes *edema* (fluid retention in tissues).

Let's get wet! Physical properties of water

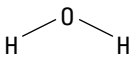
The medium in which biological systems operate is water, and physical properties of water influence the biological systems. Therefore, it is important to review some properties of water that you learned in general chemistry.

Water is a polar molecule

Because it's *polar,* water has a tendency to "wet" substances, like grandma's fine dining-room table or a baby's diaper. It's also a *bent* molecule , not linear (see Figure 2-1). The hydrogen atoms have a partially positive charge ($\delta+$); the oxygen atom has a partially negative charge ($\delta-$). This charge distribution is due to the *electronegativity* difference between hydrogen and oxygen atoms (the attraction that an atom has for a bonding pair of electrons). The water molecule in Figure 2-1 is shown in its bent shape with a bond angle of about 105°.

Normally, such partial charges result in an intermolecular force known as a *dipole-dipole force,* in which the positive end of one molecule attracts the negative end of another molecule. The very high electronegativity of oxygen combined with the fact that a hydrogen atom has only one electron results in a charge difference significantly greater than you'd normally expect. This leads to stronger-than-expected intermolecular forces. These unexpectedly strong intermolecular forces have a special name: *hydrogen bonds.*

Figure 2-1: Structure of a water molecule.



The term *hydrogen bond* doesn't refer to an actual bond to a hydrogen atom, but to the overall *interaction* of a hydrogen atom bonded to either oxygen, nitrogen, or fluorine atoms with an oxygen, nitrogen or fluorine on another molecule (intermolecular) or the same molecule (intramolecular). Hence the term *intermolecular force.* (Note that although hydrogen bonds occur when hydrogen bonds to fluorine, you don't normally find such combinations in biological systems.)

Water has strong intermolecular forces

Hydrogen bonds in oxygen- and nitrogen-containing molecules are very important in biochemistry because they influence reactions between such molecules and the structures of these biological molecules. The interaction between water and other molecules in which there may be an opportunity for hydrogen bonding explains such properties as solubility in water and reactions that occur with water as a solvent.

The term *hydrogen bond* doesn't refer to an actual bond to a hydrogen atom, but to an overall interaction.

One environmentally important consequence of hydrogen bonding is that, upon freezing, water molecules are held in a solid form that's less dense than the liquid form. The hydrogen bonds lock the water molecules into a crystalline lattice that contains large holes, which decreases the density of the ice. The less-dense ice — whether in the form of an ice cube or an iceberg — floats on liquid water. In nearly all other cases where a solid interacts with water, the reverse is true: The solid sinks in the liquid. So, why is the buoyancy of ice important? Ask ice fishermen! The layer of ice that forms on the surface of cold bodies of water insulates the liquid from the cold air, protecting the organisms still living under the ice.

Water has a high specific heat

Specific heat is the amount of heat required to change the temperature of a gram of water 1° Celsius. A high specific heat means it isn't easy to change the temperature of water. Water also has a high *heat of vaporization.* Humans can rid their bodies of a great deal of heat when their sweat evaporates from their skin, making sweat a very effective cooling method. We're sure you'll notice this cooling effect during your biochem exams.

As a result of water's high specific heat and heat of vaporization, lakes and oceans can absorb and release a large amount of heat without a dramatic change in temperature. This give and take helps moderate the earth's temperature and makes it easier for an organism to control its body temperature. Warm-blooded animals can maintain a constant temperature, and cold-blooded animals — including lawyers and some chemistry teachers — can absorb enough heat during the day to last them through the night.

Water's most important biochemical role: The solvent

The polar nature of water means that it attracts (soaks up) other polar materials. Water is often called *the universal solvent* because it dissolves so many types of substances. Many ionic substances dissolve in water, because the negative ends of the water molecules attracts the *cations* (positively charged ions) from the *ionic* compound (compound resulting from the reaction of a metal with a non-metal) and the positive ends attract the *anions* (negatively charged ions). Covalently bonded (resulting from the reactions between non-metals) polar substances, such as alcohols and sugars, also are soluble in water because of the dipole-dipole (or hydrogen-bonding) interactions. However, covalently bonded nonpolar substances, such as fats and oils are *not* soluble in water.

Polar molecules, because of their ability to interact with water molecules, are classified as *hydrophilic* (water-loving). Nonpolar molecules, which don't appreciably interact with (dissolve in) water, are classified as *hydrophobic* (water-hating). Some molecules are *amphipathic* because they have both hydrophilic and hydrophobic regions.

Figure 2-2 shows the structure of a typical amphipathic molecule. The molecule appears on the left, with its hydrophilic and hydrophobic regions shown. The alternate portion of the figure is a symbolic way of representing the molecule. The round "head" is the hydrophilic portion, and the long "tail" is the hydrophobic portion.

Certain amphipathic molecules, such as soap molecules, can form *micelles,* or very tiny droplets that surround insoluble materials. This characteristic is the basis of the cleaning power of soaps and detergents. The hydrophobic portion of the molecule (a long hydrocarbon chain) dissolves in a nonpolar substance, such as normally insoluble grease and oil, leaving the hydrophilic portion (commonly an ionic end) out in the water. Soap or detergent breaks up the grease or oil and keeps it in solution so it can go down the drain.

A micelle behaves as a large polar molecule (see Figure 2-3). The structure of a micelle is closely related to the structure of cell membranes.

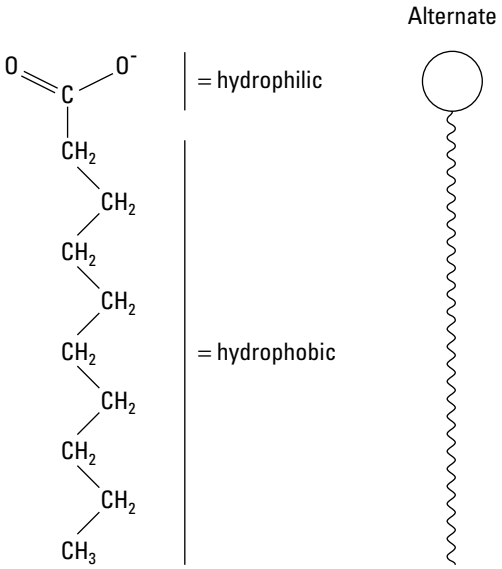


Figure 2-2: Structure of a typical amphipathic (both water-loving and water-hating) molecule.

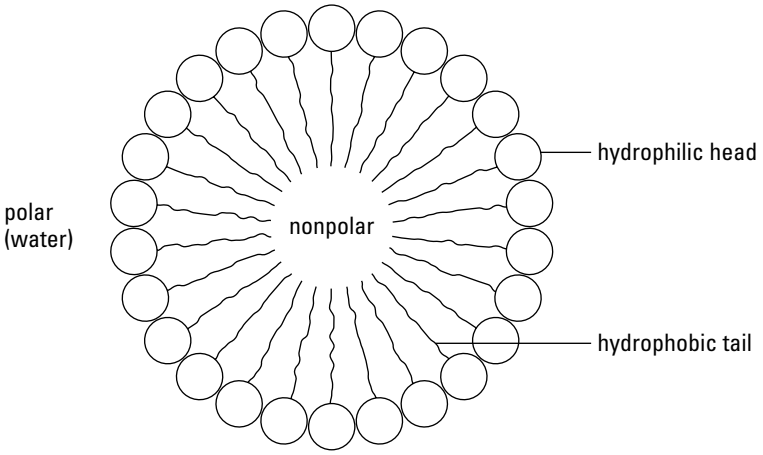


Figure 2-3: Structure of a micelle, composed of amphipathic molecules, with their hydrophilic "heads" pointing "out."

Hydrogen Ion Concentration: Acids and Bases

In aqueous solutions — especially in biological systems — the concentration of hydrogen ions (H^+) is very important. Biological systems often take great pains to make sure that their hydrogen ion concentration — represented as $[H^+]$ or by the measurement of pH (the measure of acidity in a solution) — doesn't change.

Even minor changes in hydrogen ion concentration can have dire consequences to a living organism. For example, in our blood, only a very small range of hydrogen ion allows the body to function properly. Hydrogen ion concentrations higher or lower than this range can cause death.

Because living organisms are so dependent on pH, let's take a few moments to review the concepts of acids, bases, and pH.

Achieving equilibrium

When the concentrations of hydrogen ion (H^+) and hydroxide ion (OH^-) are the same, the solution is *neutral.* If the hydrogen ion concentration exceeds the hydroxide ion concentration, the solution is *acidic.* If the hydroxide ion concentration is greater, the solution is *basic.* These chemical species are related through a chemical equilibrium.

Acidic solutions, such as lemon juice, taste sour. Basic solutions, such as tonic water, taste bitter. (The addition of gin doesn't change the bitter taste!)

The equilibrium of hydrogen ions is present in all aqueous solutions. It may or may not be the major hydrogen ion source (usually it isn't). Water is a contributor to the hydrogen ion concentration because it undergoes *autoionization*, as shown by the following equation:

$$H_2O(l) \quad H^+(aq) + OH^-(aq)$$

You often see $H^+(aq)$ represented as H_3O^+.

The double arrow ( ) indicates that this is an equilibrium; as such, there must be an associated *equilibrium constant* (K). The equilibrium constant in the preceding equation is K_w. The value of K_w is the product of the concentrations of the hydrogen ion and the hydroxide ion:

$$K_w = [H^+]\,[OH^-] = 1.0 \times 10^{-14} \text{ (at } 25^\circ C)$$

The value of the constant K_w, like all Ks, is only constant if the temperature is constant. In the human body, where $T = 37^\circ C$, $K_w = 2.4 \times 10^{-14}$.

In pure water, at $25^\circ C$, $[H^+] = 1.0 \times 10^{-7}$ M (1.6×10^{-7} M at $37^\circ C$). The hydroxide ion concentration is the same as the hydrogen ion concentration, because they are formed in equal amounts during the autoionization reaction.

M is a concentration term, the molarity. *Molarity* is the number of moles of solute per liter of solution.

Sour and bitter numbers: The pH scale

It isn't always convenient to report hydrogen ion concentrations in an exponential form, such as 1.0×10^{-7}. Thankfully, you have a way of simplifying the representation of the hydrogen ion concentration: the pH. You can calculate the pH for any solution by using the following equation:

$$pH = -\log [H^+]$$

For instance, in the case of a solution with a hydrogen ion concentration of 1.0×10^{-7} M, the pH would be

$$pH = -\log (1.0 \times 10^{-7}) = 7.0$$

Table 2-1 gives similar calculations for my hydrogen ion concentrations.

Table 2-1 The pH Scale and the Associated Hydrogen Ion Concentration

$[H^+]$	*pH*	*Solution Property*
1.0×10^{0} M	0	Acidic
1.0×10^{-1} M	1	Acidic
1.0×10^{-2} M	2	Acidic
1.0×10^{-3} M	3	Acidic
1.0×10^{-4} M	4	Acidic
1.0×10^{-5} M	5	Acidic
1.0×10^{-6} M	6	Acidic
1.0×10^{-7} M	7	**Neutral**
1.0×10^{-8} M	8	Basic
1.0×10^{-9} M	9	Basic
1.0×10^{-10} M	10	Basic
1.0×10^{-11} M	11	Basic
1.0×10^{-12} M	12	Basic
1.0×10^{-13} M	13	Basic
1.0×10^{-14} M	14	Basic

If a solution has a pH less than 7, it's acidic. Solutions with a pH greater than 7 are basic. Solutions whose pH is 7 are neutral. The pH of pure water is 7. Be careful, though: Not every solution that has a pH of 7 is pure water! For example, if you add table salt to water, the pH will remain at 7, but the resulting solution is certainly not pure water.

The pH scale is an open-ended scale, meaning you can have a pH greater than 14 or less than 0. For example, the pH of a 1.0×10^{1} M solution of hydrochloric acid is –1. John loves to ask questions based on this topic to his advanced chemistry students! The 0–14 scale is a convenient part of the pH scale for most real-world solutions — especially ones found in biochemistry. Most biological systems have a pH near 7, although significant deviations may exist (the pH in your stomach is close to 1).

Calculating pOH

You can calculate pOH in a similar manner to the pH calculation. That is, you can use the equation $pOH = -\log [OH^-]$. You can calculate the hydroxide ion concentration from the hydrogen ion concentration and the K_w (equilibrium constant) relationship:

$$[OH^-] = K_w \div [H^+]$$

A useful shortcut to get from pH to pOH is the following relationship: $pH + pOH = 14.00$ for any aqueous solution ($14.00 = pK_w = -\log K_w = -\log 1.0 \times 10^{-14}$).

For example, if a solution has a $[H^+] = 6.2 \times 10^{-6}$, its pH would be

$$pH = -\log [H^+]$$

$$pH = -\log [6.2 \times 10^{-6}]$$

$$pH = 5.21$$

The calculation for the pOH of that solution becomes pretty simple: $14.00 - pH = 14.00 - 5.21 = 8.79$.

Now, if you have the pH or pOH, getting the corresponding $[H^+]$ or $[OH^-]$ becomes a pretty simple task:

$$[H^+] = 10^{-pH} \text{ and } [OH^-] = 10^{-pOH}$$

For example, a solution with a pH of 7.35 has a $[H^+] = 10^{-7.35} = 2.2 \times 10^{-7}$.

Strong and weak: Brønsted-Lowry theory

Since the acidity (pH) of the biological medium is so very important, let's take a look at one of the most accepted theories concerning acids and bases – the Brønsted-Lowry theory. According to this theory, acids are proton (H^+) donors, and bases are proton acceptors.

Strong and weak acids

Acids increase the hydrogen ion concentration of a solution (they lower the pH, in other words). Some acids, known as *strong acids,* are very efficient at changing hydrogen ion concentration; they essentially completely ionize in water. Most acids — particularly biologically important acids — aren't very efficient at generating hydrogen ions; they only partially ionize in water. These acids are known as *weak acids.*

Bases accept (react with) rather than donate hydrogen ions in solutions. Bases decrease the hydrogen ion concentration in solutions because they react with these ions. Strong bases, although they can accept hydrogen ions very well, aren't too important in biological systems. The majority of biologically important bases are weak bases.

The Brønsted-Lowry theory helps to explain the behavior of acids and bases with respect to equilibrium. A Brønsted-Lowry acid is a hydrogen ion (H^+) donor, and a Brønsted-Lowry base is a hydrogen ion acceptor. Acetic acid, a weak acid found in vinegar, partially ionizes in solution, evidenced by the following equation:

$$CH_3COOH \quad H^+ + CH_3COO^-$$

The double arrow indicates that the acetic acid doesn't completely ionize. (For a strong acid, complete ionization would occur, and a single arrow would be present.) The equilibrium arrow ( ) indicates that all three chemical species are present in the solution: the acetic acid, the acetate ion, and the hydrogen ion, along with the water solvent.

In the Brønsted-Lowry theory, you consider the acetate ion to be a base because it can accept a hydrogen ion to become acetic acid. According to this theory, two substances differing by only one hydrogen ion — such as acetic acid and the acetate ion — are members of a *conjugate acid-base pair.* The species with one additional hydrogen ion is the *conjugate acid* (CA), and the species with one less hydrogen ion is the *conjugate base* (CB).

You can express the equilibrium from the acetate example, like all equilibria, by using a *mass-action expression* — as long as a balance among the species is present. This expression is also known as a *reaction quotient* and as an *equilibrium constant*. For acetic acid, this expression is as follows:

$$K_a = \frac{[H^+][CH_3COO^-]}{[CH_3COOH]}$$

The *a* subscript means that this expression represents an acid. The square brackets refer to the molar equilibrium concentrations of the species present. You can express the K_a as a pK_a. The calculation of pK_a is similar to the calculation of pH:

$$pK_a = -\log K_a$$

In terms of conjugate acids and bases, every K_a expression appears as

$$K_a = \frac{[H^+][CB]}{[CA]}$$

No variations are allowed in this equation other than the actual formulas of the conjugate acid and base.

Like an acid, a base has a K_b value (the subscript *b* meaning base). A weak base, like ammonia, is part of the following equilibrium:

$$NH_3 + H_2O \quad OH^- + NH_4^+$$

The equilibrium constant expression for this equilibrium is

$$K_b = \frac{[OH^-][NH_4^+]}{[NH_3]}$$

The generic form of a K_b expression is

$$K_b = \frac{[OH^-][CA]}{[CB]}$$

As with a K_a expression, a K_b expression has no variations other than the actual formulas of the conjugate acid and base.

Every conjugate acid has a K_a, and its corresponding conjugate base has a K_b. The K_a and the K_b of a conjugate acid-base pair are related by the K_w — the ionization constant for water. For a conjugate acid-base pair, $K_aK_b = K_w = 1.0 \times 10^{-14}$. In addition, you can use the following shortcut: $pK_a + pK_b = 14.00$.

The K_b for the acetate ion, the conjugate base of acetic acid, is

$$CH_3COO^- + H_2O \quad OH^- + CH_3COOH$$

The K_a for the ammonium ion, the conjugate acid of ammonia, is

$$NH_4^+ \quad H^+ + NH_3$$

An acid may be capable of donating more than one hydrogen ion. A biologically important example of this type of acid is phosphoric acid (H_3PO_4), which is a triprotic acid (meaning that it can donate three hydrogen ions). This acid is capable of donating three hydrogen ions, one at a time. The equilibria for this acid are

K_{a1}: $H_3PO_4 \quad H^+ + H_2PO_4^-$

K_{a2}: $H_2PO_4^- \quad H^+ + HPO_4^{2-}$

K_{a3}: $HPO_4^{2-} \quad H^+ + PO_4^{3-}$

The subscripts are modified to indicate the loss of hydrogen 1, hydrogen 2, or hydrogen 3. The associated K_a expressions are all of the form

$$K_a = \frac{[H^+][CB]}{[CA]}$$

Here's the breakdown for each K_a:

$$K_{a_1} = \frac{[H^+][H_2PO_4^-]}{[H_3PO_4]}$$

$$K_{a_2} = \frac{[H^+][HPO_4^{2-}]}{[H_2PO_4^-]}$$

$$K_{a_3} = \frac{[H^+][PO_4^{3-}]}{[HPO_4^{2-}]}$$

The value for each successive equilibrium constant often is significantly lower than the preceding value. Table 2-2 runs through some biologically important acids. You can refer to this table when working buffer problems or determining which acid is stronger.

Table 2-2 The K_a Values for Biologically Important Acids

Acid	K_{a1}	K_{a2}	K_{a3}
Acetic acid (CH_3COOH)	1.7×10^{-5}		
Pyruvic acid ($CH_3COCOOH$)	3.2×10^{-3}		
Lactic acid ($CH_3CHOHCOOH$)	1.4×10^{-4}		
Succinic acid ($HOOCCH_2CH_2COOH$)	6.2×10^{-5}	2.3×10^{-6}	
Carbonic acid (H_2CO_3)	4.5×10^{-7}	5.0×10^{-11}	
Citric acid ($HOOCCH_2C(OH)(COOH)CH_2COOH$)	8.1×10^{-4}	1.8×10^{-5}	3.9×10^{-6}
Phosphoric acid (H_3PO_4)	7.6×10^{-3}	6.2×10^{-8}	2.2×10^{-13}

Acid or base? They just can't decide

Some substances can't make up their minds about what they are; they can act as either an acid or a base. Chemists classify these substances as *amphiprotic* or *amphoteric* substances. For example, the bicarbonate ion (HCO_3^-) can act as either an acid or a base:

$$HCO_3^- \quad H^+ + CO_3^{2-}$$

$$HCO_3^- + H_2O \quad OH^- + H_2CO_3$$

Biochemically important molecules may also exhibit amphiprotic behavior. Amino acids contain both a basic amine ($-NH_2$) group and an acidic carboxyl ($-COOH$) group. Therefore, they can act as either acids or bases. For example, glycine (H_2N-CH_2-COOH) may undergo the following reactions:

$$H_2N-CH_2-COOH \quad H^+ + H_2N-CH_2-COO^-$$

$$H_2N-CH_2-COOH + H_2O \quad OH^- + {}^+H_3N-CH_2-COOH$$

In fact, amino acids may undergo proton transfer from the carboxyl end to the amine end, forming an overall neutral species that has a positive and negative end. Species such as these are called *zwitterions:*

$H_2N–CH_2–COOH \quad {}^{+}H_3N–CH_2–COO^{-}$

Buffers and pH Control

A solution that contains the conjugate acid-base pair of any weak acid or base is a *buffer solution.* A buffer solution resists changes in pH when either an acid or a base is added. Therefore, buffers control the pH of the solution. Buffer solutions are important in most biological systems. Many biological processes proceed effectively only within a limited pH range. The presence of buffer systems keeps the pH within this limited range.

Identifying common physiological buffers

In the human body, the pH of various body fluids is important. The pH of blood is 7.4, the pH of stomach acid is 1–2, and the pH in the intestinal tract is 8–9. If the pH of blood is more than 0.2 pH units lower than normal, a condition known as *acidosis* results; a corresponding increase in pH is *alkalosis.* Acidosis and alkalosis may lead to serious health problems. There are two general causes of acidosis and of alkalosis:

- Respiratory acidosis is the result of many diseases that impair respiration. These diseases include pneumonia, emphysema, and asthma. These diseases lead to inefficient expulsion of carbon dioxide. This leads to an increase in the concentration of the acid H_2CO_3.
- Metabolic acidosis is due to a decrease in the concentration of HCO_3^-. This may be the results of certain kidney diseases, uncontrolled diabetes, and cases of vomiting involving nonacid fluids. Poisoning by an acid salt may also lead to metabolic acidosis.
- Respiratory alkalosis may be the result of hyperventilation, as there is an excessive removal of carbon dioxide, which leads to a decrease in the H_2CO_3 concentration. Immediate treatment may include breathing into a paper bag, thus increasing the carbon dioxide concentration in the inhaled air and therefore in the blood.
- Metabolic alkalosis may result from excessive vomiting of stomach acid.

To resist these pH problems, the blood has a number of buffer systems. These include several proteins in blood plasma and the bicarbonate buffer system.

The *bicarbonate buffer system* is the main extracellular buffer system. This system also provides a means of eliminating carbon dioxide. The dissolution of carbon dioxide in aqueous systems sets up the following equilibria:

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

The presence of the conjugate acid-base pair (H_2CO_3 and HCO_3^-) means this is a buffer system. The conjugate acid-base ratio is about 20:1 at a pH of 7.4 in the bloodstream. This buffer system is coupled with the following equilibrium (nstrumental in the removal of carbon dioxide in the lungs):

$$CO_2(blood) \quad CO_2(g)$$

The second ionization of phosphoric acid, K_{a2}, is the primary intracellular buffer system. The pH of this conjugate acid-base pair ($H_2PO_4^-$ and HPO_4^{2-}) is 7.21 for a solution with equal concentrations of these two species.

Calculating a buffer's pH

To determine a buffer's pH, you may use a K_a or K_b calculation, discussed previously, or the Henderson-Hasselbalch equation, which gives a shortcut.

There are two forms of the Henderson-Hasselbalch equation:

$$pH = pK_a + \log\frac{[CB]}{[CA]}$$

and

$$pOH = pK_b + \log\frac{[CA]}{[CB]}$$

The terms in either form are the same as defined earlier. For example, suppose we wanted to calculate the pH of a buffer composed of 0.15 M pyruvic acid and 0.25 M sodium pyruvate. Looking at Table 2-2, you see that the K_a of pyruvic acid is 3.2×10^{-3}.

The pK_a would be 2.50. Therefore:

$$pH = pK_a + \log\frac{[CB]}{[CA]}$$

$$pH = -\log 3.2 \times 10^{-3} + \log\frac{[CH_3COCOO^-]}{[CH_3COCOOH]}$$

$$pH = 2.50 + \log\frac{[.25]}{[.15]}$$

$$pH = 2.50 + \log(1.67)$$

$$pH = 2.50 + 0.22 = 2.72$$

The greater the values of [CA] and [CB], the greater the buffer capacity of the solution is. The buffer capacity indicates how much acid or base may be added to a buffer before the buffer ceases to function.

Chapter 3

Fun with Carbon: Organic Chemistry

In This Chapter

- Understanding why carbon is fundamental to biochemistry
- Examining the nature of weak interactions
- Finding out about functional groups
- Checking out isomerism

Most biologically important molecules are composed of *organic compounds,* which means compounds of carbon. Therefore, the student of biochemistry must have a general knowledge of organic chemistry, which is the study of carbon compounds, in order to understand the function and reactions of biochemical molecules. In this chapter we review the basics of organic chemistry, including the various functional groups and isomers that are important in the field of biochemistry.

The Role of Carbon in the Study of Life

Long ago, scientists believed that all carbon compounds were the result of biological processes, which meant organic chemistry was synonymous with biochemistry under this Vital Force theory. In the mid-1800s, though, researchers debunked that long-held notion — the synthesis of urea from inorganic materials showed that there were other paths to the production of carbon compounds. Organic chemists now synthesize many important organic chemicals without the use of living organisms; however, biosynthesis is still an important source of many organic compounds.

Why are there so many carbon compounds? The answer lies primarily in two reasons, both tied to carbon's versatility in creating stable bonds:

- **Carbon bonds to itself.** Carbon atoms are capable of forming stable bonds to other carbon atoms. The process of one type of atom bonding to identical atoms is *catenation.* Many other elements can catenate, but carbon is the most effective at it. There appears to be no limit to how many carbon atoms can link together. These linkages may be in chains, branched chains, or rings, as shown in Figure 3-1.
- **Carbon bonds to other elements.** Carbon is capable of forming stable bonds to a number of other elements. These include the biochemically important elements hydrogen, nitrogen, oxygen, and sulfur. The latter three elements form the foundation of most of the *functional groups* (reactive groups of a molecule) necessary for life. Bonds between carbon and hydrogen are usually unreactive under biochemical conditions; thus, hydrogen often serves as an "inert" substituent.

It's All in the Numbers: Carbon Bonds

Carbon is capable of forming four bonds. In bonding to itself and other elements, carbon uses a variety of types of *hybridization* — when it bonds to another carbon molecule, for example, these may have four single bonds, a double and two single bonds, two double bonds, or a triple and a single bond. Double bonds to oxygen atoms are particularly important in many biochemicals. Table 3-1 shows the number of bonds carbon may have with some selected non-metals, along with the hybridization of those bonds.

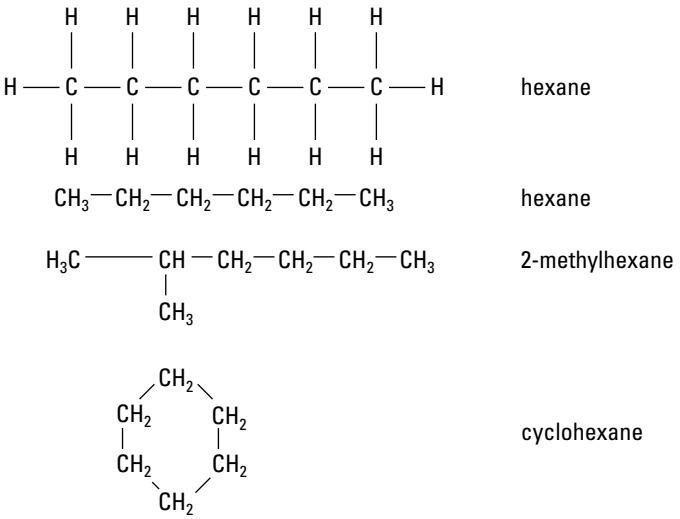


Figure 3-1: Top: straight chain hydrocarbon, expanded and condensed. Middle: branched chain hydrocarbon. Bottom: ring hydrocarbon.

Table 3-1 Possible Bonds of Carbon and Selected Non-metals

Element	*Number of Possible Bonds with Carbon*	*Some Possible Hybridizations for Second Period Elements*
Carbon (C)	4	4 single (sp^3) 2 single and one double (sp^2) 1 single and one triple (sp) 2 doubles (sp)
Nitrogen (N)	3	3 single (sp^3) 1 single and 1 double (sp^2) 1 triple (sp)
Oxygen (O)	2	2 single (sp^3) 1 double (sp^2)
Sulfur (S)	2	2 single (sp^3) 1 double (sp^2)
Hydrogen (H)	1	1 single
Fluorine (F)	1	1 single
Chlorine (Cl)	1	1 single
Bromine (Br)	1	1 single
Iodine (I)	1	1 single

Sticky Chemistry: Bond Strengths

Covalent bonds are important *intramolecular forces* (forces within the same molecule) in biochemistry. *Intermolecular forces* (forces between chemical species) are also extremely important. Among other things, intermolecular forces are important to hydrophilic (water-loving) and hydrophobic (water-hating) interactions.

Everybody has 'em: Intermolecular forces

All intermolecular forces are *van der Waals forces,* that is they are not true bonds in the sense of sharing or transferring electrons, but are weaker attractive forces. These forces include dipole-dipole forces, hydrogen bonding, and ionic interactions.

Dipole-dipole forces

Dipole-dipole forces exist between polar regions of different molecules. The presence of a dipole means that the molecule has a partially positive ($\delta+$) end and a partially negative ($\delta-$) end. Opposite partial charges will attract each other, whereas like partial charges will repel. In most cases, biological systems utilize a special type of dipole-dipole force known as *hydrogen bonding* (see next section).

Hydrogen bonding

Hydrogen bonding, as the name implies, involves hydrogen. The hydrogen atom must be bonded to either an oxygen atom or a nitrogen atom. (In non-biological situations, hydrogen bonding also occurs when a hydrogen atom bonds to a fluorine atom.) Hydrogen bonding is significantly stronger than a "normal" dipole-dipole force and is very much stronger than London dispersion forces (very weak and short-lived attractions between molecules that arise due to the nucleus of one atom attracting the electron cloud of another atom). The hydrogen that is bonded to either a nitrogen or oxygen atom is strongly attracted to a different nitrogen or oxygen atom. Hydrogen bonding may be either intramolecular or intermolecular.

Ionic interactions

In biological systems, ionic interactions may serve as intermolecular or intramolecular forces. In some cases, these may involve metal cations, such as Na^+, or anions, such as Cl^-. In many cases, the cation is an ammonium ion from an amino group, such as RNH_3^+; the anion may be from a carboxylic acid, such as $RCOO^-$. Oppositely charged ions attract each other strongly.

Water-related interactions: Both the lovers and the haters

The predominant factor leading to hydrophobic (water-hating) interactions is the presence of portions of a molecule containing only carbon and hydrogen. Hydrocarbon regions are nonpolar and are attracted to other nonpolar regions by London dispersion forces.

In general, the presence of any atom other than carbon and hydrogen makes a region polar. Oxygen and nitrogen are the most effective elements in biochemistry for making a region of a molecule polar. Sulfur is least effective of the common biologically important elements at imparting polar character. Dipole-dipole, hydrogen bonding, and ionic interactions are all hydrophilic interactions. London dispersion forces are hydrophobic interactions.

The more carbon and hydrogen atoms, without other atoms, the more important the hydrophobic nature of a region becomes in defining the properties of the molecule. Note that a molecule may have both a hydrophilic and a hydrophobic region, and both regions are important to the behavior of the molecule. The formation of a micelle is an example of using molecules with both hydrophilic and hydrophobic regions.

How bond strengths affect physical properties of substances

The physical properties of biological substances depend on the intermolecular forces present. The sequence of strength is: ions > hydrogen bonding > dipole-dipole > London. The strongest types of intermolecular forces involve ions. Next strongest is hydrogen bonding. Polar substances interact through dipole-dipole forces, which are weaker than hydrogen bonds. All biological substances containing oxygen, nitrogen, sulfur, or phosphorus are polar. London forces, the weakest intermolecular forces, are important in nonpolar situations. The hydrocarbon portion of biological molecules is nonpolar.

Melting points, boiling points, and solubility

As the strength of forces decreases, so do the melting points, boiling points, and solubility in water. Besides that, the vapor pressure and the solubility in nonpolar solvents also increases.

Substances that have a high solubility in water are hydrophilic, and substances with a low solubility in water are hydrophobic.

A molecule may have both hydrophilic and hydrophobic regions. The region that represents a greater portion of the molecule will predominate. For this reason, for example, CH_3COOH is more hydrophilic than $CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2COOH$. In addition, $HOCH_2CH_2CH_2CH_2CH_2CH_2OH$ is more hydrophilic than $CH_3CH_2CH_2CH_2CH_2CH_2OH$.

Odors

Many functional groups have distinctive odors. Small carboxylic acids smell like acetic acid (vinegar), while larger ones have unpleasant odors. Most esters, if volatile, have pleasant odors — that is why esters are used extensively in the flavor and perfume industry. Most sulfur-containing compounds have strong unpleasant odors. Small amines have an ammonia odor, whereas larger amines have a fishy odor or worse.

Defining a Molecule's Reactivity: Functional Groups

Most carbon compounds have one or more reactive sites composed of a specific grouping of atoms in their structure. It is at these sites that chemical reactions occur. These specific grouping of atoms that react are called *functional groups.* These functional groups contain atoms other than carbon and hydrogen and/or double or triple bonds and define the reactivity of the organic molecule.

Hydrocarbons

Alkanes are hydrocarbons — compounds containing only carbon and hydrogen, with no traditional functional groups. For this reason, they are not very reactive. *Alkenes* and *alkynes* are also hydrocarbons. They contain a carbon-carbon double and triple bond, respectively. The presence of more than one bond makes them more reactive. *Aromatic hydrocarbons,* normally ring structures with alternating single and double carbon-to-carbon bonds, contain one or more *aromatic systems*, which are much less reactive than other systems containing double bonds. Alkynes are not very common in biological systems. Figure 3-2 shows the structure of these compounds.

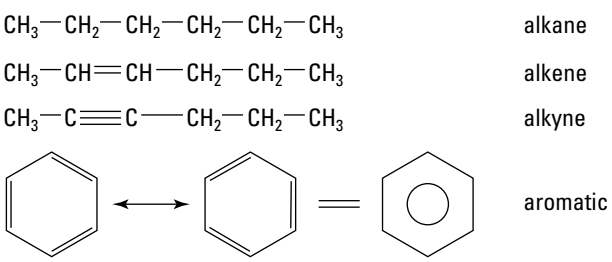


Figure 3-2: Examples of alkanes, alkenes, alkynes, and aromatic hydrocarbons.

Functional groups with oxygen and sulfur

Many functional groups contain oxygen, including alcohols, ethers, aldehydes, and ketones, which appear in carbohydrates. In carbohydrates, many ether groups are known as *glycoside linkages*. In addition, carboxylic acids and esters are important functional groups that appear as fatty acids and in fats and oils.

Alcohols and ethers contain only singly bonded oxygen atoms. An alcohol group attached to an aromatic ring is a phenol. Aldehydes and ketones contain only doubly bonded oxygen atoms. Carboxylic acids and esters contain both singly and doubly bonded oxygen atoms. The combination of a carbon atom connected to an oxygen atom by a double bond is a carbonyl group.

Sulfur, the element immediately below oxygen on the periodic table, may replace oxygen in both alcohols and ethers to give *thiols* (mercaptans) and *thioethers.* Many of these sulfur-containing compounds really stink! Sulfur may also form a disulfide where there is a bond between two sulfur atoms. Figure 3-3 illustrates these compounds.

R = any organic (hydrocarbon) group

R′ = any organic group, which may or may not = R

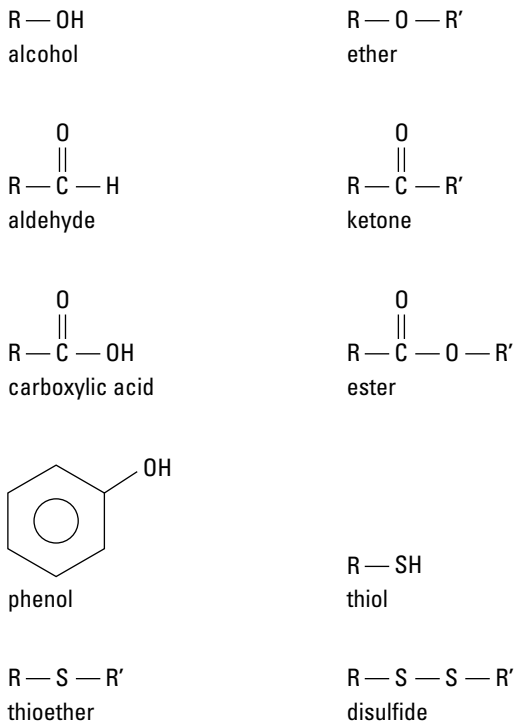


Figure 3-3: Oxygen- and sulfur-containing functional groups.

Functional groups containing nitrogen

Amines and amides are two important functional groups containing nitrogen. *Amines* are present in amino acids and alkaloids. *Amides* are present in proteins, where they are known as *peptide bonds.*

The difference between an amine and an amide is that amides have a carbonyl group adjacent to the nitrogen atom. Amines are derivatives of ammonia, NH_3, where one or more organic groups replace hydrogen atoms. In a *primary amine,* an organic group replaces one hydrogen atom. In *secondary* and *tertiary amines,* two and three organic groups, respectively, replace two or three hydrogen atoms. Figure 3-4 shows these compounds.

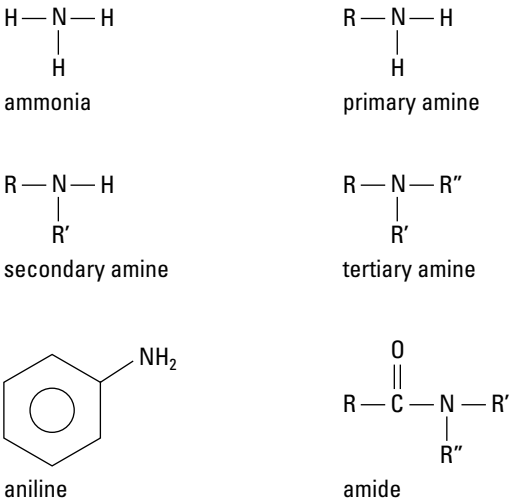


Figure 3-4: Some nitrogen-containing functional groups.

Alkaloids are basic compounds produced by plants. Examples include nicotine, caffeine, and morphine.

Functional groups containing phosphorus

Phosphorus is also an important element in biological systems and is normally present as part of a phosphate group. Phosphate groups come from phosphoric acid, H_3PO_4. The phosphate groups may be alone, part of a diphosphate, part of a triphosphate, or part of a phosphate ester.

Phosphates appear in teeth and bone and are a part of the energy transport molecules ATP and ADP (see Chapter 12 for more on these). Figure 3-5 illustrates phosphorous-containing functional groups.

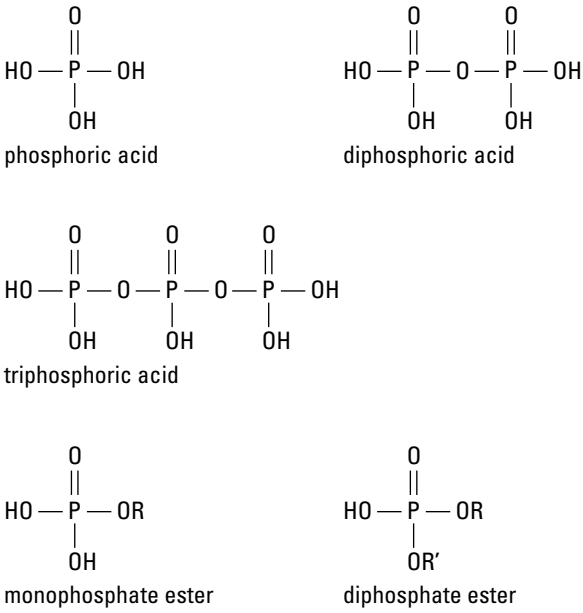


Figure 3-5: Phosphate-containing functional groups.

Reactions of functional groups

As you study the different biochemical molecules and their functions within the living organism, you see that the way a certain molecule reacts is primarily determined by the functional groups in the molecule's structure. Take a few minutes and refresh your organic chemistry knowledge about the typical reactions of the various functional groups.

Alcohols

Alcohols are subject to oxidation. Mild oxidation of a primary alcohol (where the –OH is attached to an end carbon) produces an aldehyde, which may undergo further oxidation to a carboxylic acid. Under similar conditions, a secondary alcohol (–OH is attached to a carbon bonded to two other carbons) will yield a ketone, and a tertiary alcohol (–OH attached to a carbon bonded to three other carbons) will not react. This behavior is important in the chemistry of many carbohydrates.

The presence of the OH leads people mistakenly to assume that alcohols are bases. Nothing could be further from the truth! Alcohols, under biological conditions, are neutral compounds. Phenols, though, are weak acids.

Aldehydes and ketones

Aldehydes easily undergo oxidation to carboxylic acids, but ketones do not undergo mild oxidation. With difficulty, it is possible to reduce aldehydes and ketones back to the appropriate alcohols.

Reducing sugars behave as such due to mild oxidation of the carbonyl groups present. Tollen's test uses silver nitrate, in which a reducing sugar generates a silver mirror on the walls of the container. Both Benedict's and Fehling's tests use copper compounds, and a reducing sugar produces a red precipitate with either of these tests. These simple organic qualitative tests find some use in biochemical tests described later in this book.

The carbonyl group of an aldehyde or ketone may interact with an alcohol to form acetals and hemiacetals. (Modern terminology only uses the terms acetals and hemiacetals — you may sometime see the terms *hemiketal,* which is a type of hemiacetal, and *ketal,* a type of acetal.) See Figure 3-6 for an illustration of these.

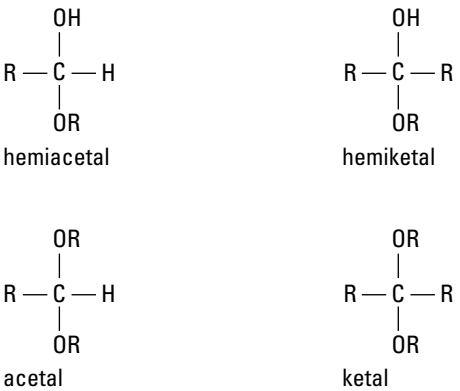


Figure 3-6: Acetals, hemiacetals, hemiketals, and ketals.

Carboxylic acids

Carboxylic acids, along with phosphoric acid, are the most important biological acids. Carboxylic acids react with bases such as the amines to produce salts. The salts contain an ammonium ion from the amine and a carboxylate ion from the acid.

Carboxylic acids combine with alcohols to form esters and can indirectly combine with amines to form amides. Hydrolysis of an ester or an amide breaks the bond and inserts water. An acid, base, or enzyme is needed to

catalyze it. Under acidic conditions, it is possible to isolate the acid and either the alcohol or the ammonium ion from the amine. Under basic conditions, you can isolate the carboxylate ion and either the alcohol or the amine.

Thiols and amines

Under mild oxidation, two thiols join to form a disulfide. Mild reducing conditions, catalyzed by enzymes, reverse this process. Such formation of disulfide linkages is important in the chemistry of many proteins, such as insulin.

Amines are the most important biological bases. As bases, they can react with acids. The behavior is related to the behavior of ammonia.

$NH_3 + H^+$ (from an acid) $\rightarrow NH_4^+$ (ammonium ion)

$NRH_2 + H^+$ (from an acid) $\rightarrow NRH_3^+$ (ammonium ion)

$NR_2H + H^+$ (from an acid) $\rightarrow NR_2H_2^+$ (ammonium ion)

$NR_3 + H^+$ (from an acid) $\rightarrow NR_3H^+$ (ammonium ion)

Many medications have amine groups. In order to make them more readily soluble, it is necessary to convert many of these amines to ammonium ions. For example, the reaction of the medication with hydrochloric acid forms the chloride, which often appears on the label as the hydrochloride.

It is possible to replace all the hydrogen atoms from an ammonium ion, NH_4^+, to produce a quaternary ammonium ion, NR_4^+.

Phosphoric acid

Phosphoric acid, H_3PO_4, may behave like a carboxylic acid and form esters. The esters will have an organic group, R, replacing one, two, or three of the hydrogen atoms. The resultant compounds are monoesters, diesters, and triesters. The hydrogen atoms remaining in the mono- and diesters are acidic.

pH and functional groups

Many of the biological functions of substances are pH dependent. For this reason, it is important to know which functional groups are acidic, basic, or neutral. Neutral functional groups behave the same no matter what the pH is. Table 3-3 lists the functional groups and whether or not they are acidic, neutral, or basic. In addition, we list whether they are of medium weakness, weak, or very weak. The weaker a substance in terms of pH, the less likely it will be affected by its solution pH.

Table 3-3 Acid-Base Properties of Biologically Important Functional Groups

Acids:		
	Monophosphate esters and diphosphate esters (medium)	
	Carboxylic acids (weak)	
	Phenols (very weak)	
	Thiols (very weak)	
	Amine salts (very weak)	
Bases:		
	Amines (weak)	
	Carboxylate ions (very weak)	
Neutral:		
	Alcohols	Carboxylic esters
	Ethers	Triphosphate esters
	Thioethers	Disulfides
	Amides	Ketones
	Aldehydes	

Same Content, Different Structure: Isomerism

Isomers are compounds that have the same molecular formula, but different structural formulas. Some organic and biochemical compounds may exist in different isomeric forms. Many times, especially in biological systems, these different isomers have different properties. The two most common types of isomers in biological systems are cis-trans isomers and isomerism due to the presence of a chiral carbon.

Cis-trans isomers

The presence of carbon-carbon double bonds leads to the possibility of having isomers present. Double bonds are rather restrictive and limit molecular movement. Groups on the same side of the double bond tend to remain in that position (cis), whereas groups on opposite sides tend to remain across the bond from each other (trans). See Figure 3-7 for an illustration.

Figure 3-7: Cis and trans isomers.

If the two groups attached to either of the carbon atoms of the double bond are the same, cis-trans isomers are not possible. Cis isomers are the normal form of fatty acids, whereas food processing tends to convert some of the cis isomers to the trans isomers.

Cis-trans isomers are also possible in cyclic systems. The cis form has similar groups on the same side of the ring, whereas the trans form has similar groups above and below the ring.

Chiral carbons

Trying to put your gloves on the wrong hands is kind of like another property of biological systems: handedness. There are *left-handed* molecules and *right-handed* molecules.

Identifying chiral molecules

The presence of an asymmetric, or *chiral,* carbon atom is sufficient to produce a "handed" molecule.

A chiral carbon atom has four different groups attached to it. The majority of biological molecules have one or more chiral carbon atoms and, for this reason, they are chiral. Figure 3-8 shows the chiral nature of glucose.

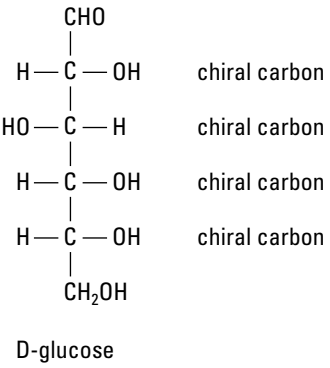


Figure 3-8: The structure of glucose, a sugar with four chiral carbon atoms.

Determining the chiral form: Enantiomer or sterioisomer?

All substances have a mirror image (okay, except vampires); however, if there is a chiral carbon atom present, the mirror images are nonsuperimposable. Hold out your left and right hands, palms up — they are nonsuperimposable mirror images. These two mirror images are called *enantiomers.* The different chiral forms differ from each other in two aspects:

- How they affect light
- How they interact with other chiral substances (usually only one chiral form will be biologically active)

To determine how a particular form affects light, it is necessary to use *plane polarized light,* in which all the light waves vibrate in the same plane. When you use this kind of light, a chiral substance rotates the vibrational plane of the light — one form (the dextrorotatory, *d*, (+) isomer) rotates the plane to the right, while the other (the levorotatory, *l*, (–) isomer) rotates the plane to the left. The *d* and *l* forms are *stereoisomers* and are optically active.

Illustrating the chiral compound: Fisher projection formulas

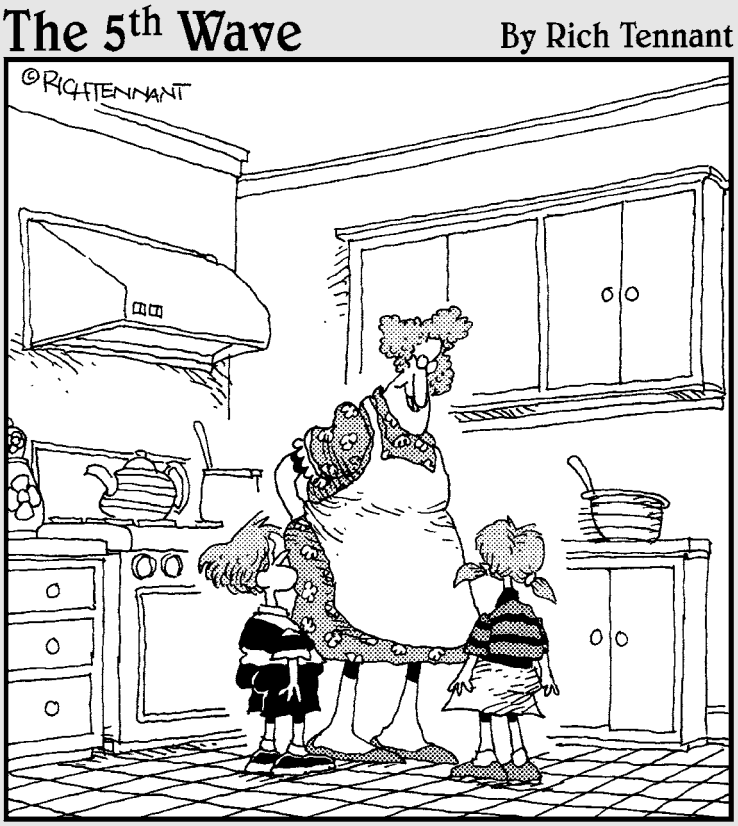
A chemist named Emil Fischer developed a method of drawing a compound to illustrate which stereoisomer was present. These Fischer projection formulas are very useful in biochemistry. In a projection formula, a chiral carbon is placed in the center of a + pattern. The vertical lines (bonds) are pointing away from the viewer, and the horizontal lines are pointing toward the viewer. Fischer used the D designation if the most important group (the group whose central atom had the largest atomic number) was to the right of the carbon, and the L designation if the most important group (lowest atomic number) was to the left of the carbon. Figure 3-9 shows two Fischer projection formulas.

Figure 3-9: Fischer projection formulas distinguish stereoisomers.

The *d* and *l* symbols are not necessarily the D and L forms respectively; thus, confusion may occur and lead to incorrect predictions. For this reason, the use of *d* and *l* is diminishing. The use of D and L is gradually being replaced by the R and S system of designating isomers. This system is particularly useful when more than one chiral carbon atom is present. For a description of this system, see *Organic Chemistry For Dummies* by Arthur Winter (Wiley).

Part II

The Meat of Biochemistry: Proteins



"Who wants to help Grandma make her famous gingerbread man cookies? You kids get the flour, eggs, and sugar, and I'll get the amino acids and enzymes."

In this part . . .

We focus, not surprisingly, on proteins, starting with amino acids, protein's building blocks. After that we detail the processes of amino acid sequencing and the various kinds of protein structure. We finish up this part by discussing enzyme kinetics, covering catalysts (which speed up reactions) and inhibitors (which — can you guess? That's right — slow them down).

Chapter 4

Amino Acids: The Building Blocks of Protein

In This Chapter

- Looking at the structure and properties of amino acids
- Examining the common amino acids
- Finding out about the interactions of amino acids
- Seeing how amino acids combine

All cells contain thousands of types of proteins, and amino acids are the building blocks of these proteins. The sequential order, number, and chemical identity of the amino acids in the protein determine the structure of the protein as well as how the protein functions. That's why it's important to understand the chemical properties of amino acids before you can understand the behavior of proteins.

Amino acids are relatively simple molecules containing both an amine group and an acid group. The biologically important amino acids are the α-amino acids that have the amine and acid groups attached to the same carbon atom. There are more than 100 known natural amino acids; however, only 20 of them are used in protein synthesis. Francis Crick (who with James Watson determined the structure of DNA) labeled this set of amino acids the *magic 20*. Other amino acids are found in certain proteins, but in almost all cases these additional amino acids result from the modification of one of the magic 20 after the protein formed.

In this chapter, we examine the structure and properties of amino acids, especially the more common ones, and show how they interact and combine.

General Properties of Amino Acids

Like any organic compound, the properties of the molecules are largely determined by the functional groups present. In biological systems, the important properties of the amino acids include the following:

- **They can join to form proteins.** The average molecular weight of an amino acid is about 135. Proteins have molecular weights ranging from about 6,000 to several million. Thus, a large number of amino acids must be joined together to produce a protein.
- **They all have both an acid and a base.** The α-carbon (end carbon) not only has an amine group ($-NH_2$) and a carboxylic acid group (–COOH), but also two additional groups: a hydrogen atom and an R– group. The side chain, R group, identifies the amino acid.
- **They all have variations in what part of the structure is protonated depending on the pH of the solution and the structure of the rest of the molecule.**
- **They all, except glycine, have a chiral nature, influencing the reactions that the compound will undergo.**

Amino acids are positive and negative: The zwitterion formation

The presence of both an acid and a base (amine) in the same molecule leads to an interaction between the two. This interaction results in a transfer of a hydrogen ion from the acid portion to the base portion of the molecule. An amino acid with both positive and negative regions is a called a *zwitterion.* The net charge of the zwitterion is 0. This leaves the acid end of the amino acid with a negative charge ($-COO^-$) and a positive charge at the base end ($-NH_3^+$). The deprotonated portion (portion that has lost a hydrogen ion) is a carboxylate group, and the protonated group (group that has gained a hydrogen ion) is an ammonium group. The presence of a charge on the amino acid makes them water-soluble. Figure 4-1 shows zwitterion formation.

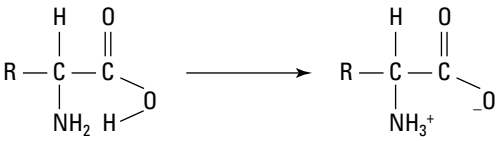


Figure 4-1: Zwitterion formation.

The unionized amino acid molecule shown in Figure 4-1 does not actually exist. However, many books and instructors draw the unionized form as a simplification, as if the ionization did not occur.

Protonated? pH and the isoelectric point

How amino acids react, because of their acid-base nature, is dependent on the pH of the solution in which they are found. Here we look at some of the implications of this pH dependency. The zwitterion is the predominant form at a particular pH, which is designated the *isoelectric point* (pI). The isoelectric point is midway between the two different pK_a values. Under most physiological conditions, isolated amino acids exist in their zwitterion form (Figure 4-2 (a)). Pure amino acids are also in the zwitterion form — and, for this reason, are *ionic solids.*

- **At a pH below the isoelectric point, some of the carboxylate groups will be protonated.** (See Figure 4-2 (b).) The pH required to cause this protonation of the carboxylate group depends on the K_a of the acid. For this reason the pK_a of the carboxylic acid group is important. Typical values are between 1 and 3. If, for example, the pK_a is 2.5, at a pH of 2.5, 50 percent of the carboxylate groups will be protonated. The net charge of the protonated form is +1.
- **At a pH above the isoelectric point, some of the ammonium groups will be deprotonated.** (See Figure 4-2 (c).) The pH required to cause this deprotonation of the ammonium group depends on the K_a of the ammonium group. For this reason, the pK_a of the ammonium group is important. Typical values are between 8 and 11. If, for example, the pK_a is 10, at a pH of 10, 50 percent of the ammonium groups will be deprotonated. The net charge of the protonated form is –1.

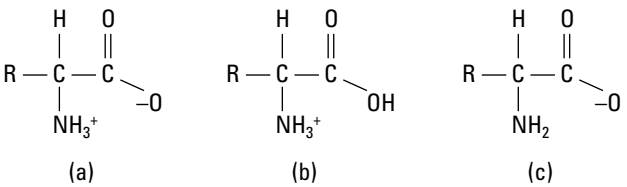


Figure 4-2: (a) Zwitterion form, (b) protonated form, and (c) deprotonated form.

Some of the side chains are also acidic or basic. In these cases, an additional pK_a becomes significant in the reactions of these molecules and will obviously complicate the pH behavior of the amino acid.

Asymmetry: Chiral amino acids

In a typical α-amino acid, four different groups are attached to the α-carbon (–COOH, $–NH_2$, –R, and –H). This makes the α-carbon asymmetric or *chiral*. The only exception is the amino acid glycine, where the R– group is a hydrogen atom. The presence of two hydrogen atoms on the α-carbon means that, in the case of glycine, the carbon atom is achiral. Chiral materials are optically active; the different forms affect light in different ways. (See Chapter 3 for more on what makes a molecule chiral.)

The arrangement of the groups around a chiral carbon atom is important. Just as your left hand only fits into your left glove, only certain arrangements of the groups will fit (because of what is called *handedness*).

There are two different forms of the chiral amino acids: the D- and the L- forms. Only the L- forms are constituents of proteins. The D- forms appear in some antibiotics and in the cell walls of certain bacteria. Fischer projections, as we explain in Chapter 3, are commonly used to represent the arrangement about the chiral carbon. Figure 4-3 illustrates some different ways to draw the Fischer projections of the structure of amino acids.

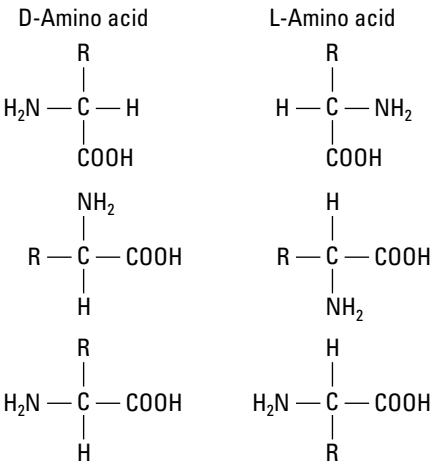


Figure 4-3: Different ways of drawing the Fischer projections of the amino acids.

A few amino acids contain two asymmetric carbon atoms. In these cases, there are four possible isomers. Biological activity is usually limited to only one of these four isomers.

The Magic 20 Amino Acids

Amino acids are subdivided into four subgroups based on the nature of the side chain (groups attached to the the α-carbon) and the general behavior of the amino acid:

- Nonpolar (hydrophobic) and uncharged
- Polar (hydrophilic) and uncharged
- Acidic (polar and charged)
- Basic (polar and charged)

The properties of the side chains are not only important to the behavior of the individual amino acids but also to the properties of the proteins resulting from the combination of certain amino acids.

In the following section we examine the structures of the individual amino acids. It is possible to represent each of the amino acids by either a three-letter or a one-letter abbreviation. Like the chemical symbols for the elements, these are fixed abbreviations. The three-letter abbreviations are easier to relate to the name of the specific amino acid. For example, we use glu for glutamine. The one-letter abbreviations are shorter, but not always related to the name. For example, we use Q for glutamine.

Nonpolar (hydrophobic) amino acids

The nonpolar amino acids are as follows:

- Alanine (ala, A)
- Valine (val, V)
- Leucine (leu, L)
- Isoleucine (ile, I)
- Proline (pro, P)
- Methionine (met, M)
- Phenylalanine (phe, F)
- Tryptophan (trp, W)

Figure 4-4 shows these amino acids.

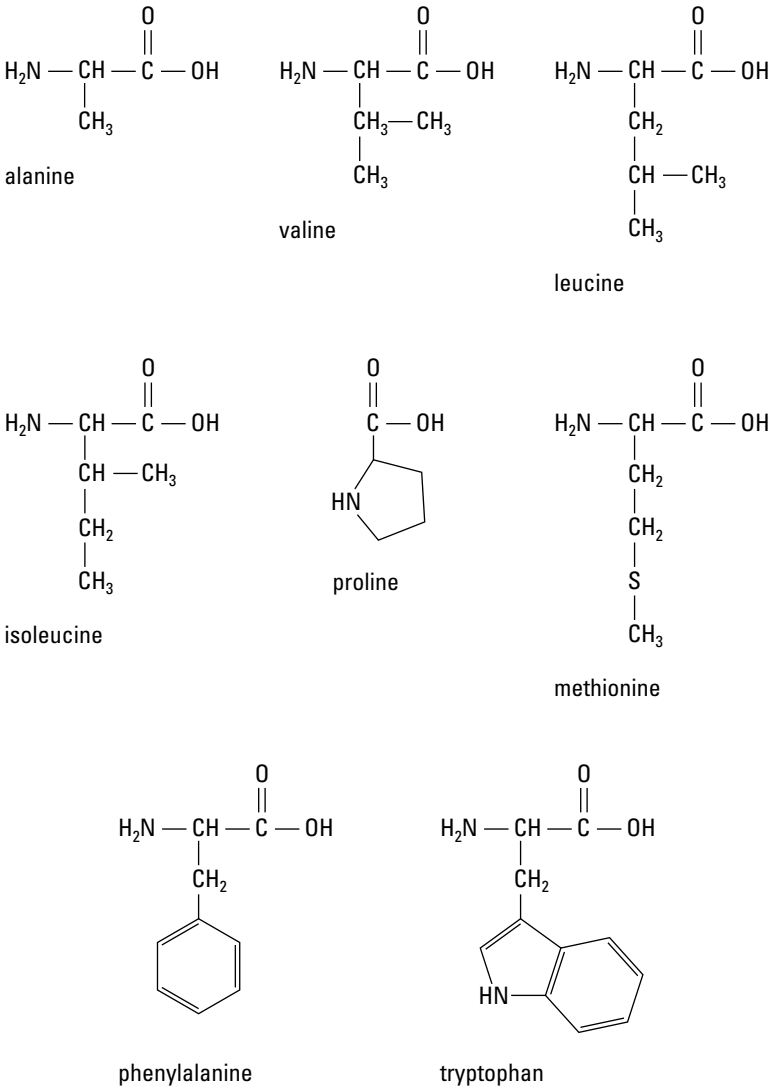


Figure 4-4: Nonpolar amino acids.

Proline has an unusual cyclic structure, which has a significant influence on protein structure. Tryptophan is a borderline case because the –NH from the ring system can interact with water to a limited extent.

Polar and uncharged (hydrophilic) amino acids

The polar and uncharged amino acids, other than glycine, can hydrogen bond to water. For this reason, they are usually more soluble than the nonpolar amino acids. The amino acids in this group are as follows:

- Glycine (gly, G)
- Serine (ser, S)
- Asparagine (asn, N)
- Glutamine (gln, Q)
- Threonine (thr, T)
- Tyrosine (tyr, Y)
- Cysteine (cys, C)

Glycine seems to be an unexpected member of this group. The small size of the R group in the case of glycine leads to the predominance of the amino and carboxylate functional groups, giving glycine's similarity to other amino acids in this group. The amide, alcohol, and sulfhydryl (–SH) groups of the remaining members of this group are very polar and neutral. At very high pH values, the phenolic group on tyrosine ionizes to yield a polar charged group. Figure 4-5 shows these amino acids.

Acidic amino acids

The acidic amino acids are as follows:

- Aspartic acid (asp, D)
- Glutamic acid (glu, E)

In both of these amino acids, the side group contains a carboxylic acid group. This secondary carboxylic acid group is a weaker acid (higher pK_a) than the primary carboxylic acid group. This additional carboxylate group leads to a net –1 charge at a pH where the "normal" zwitterion has a 0 net charge. The carboxylate side chain is important in the interaction of many proteins with metal ions, as *nucleophiles* (an electron-rich group replacing some group attached to a carbon) in many enzymes, and in ionic interactions. Figure 4-6 shows these amino acids.

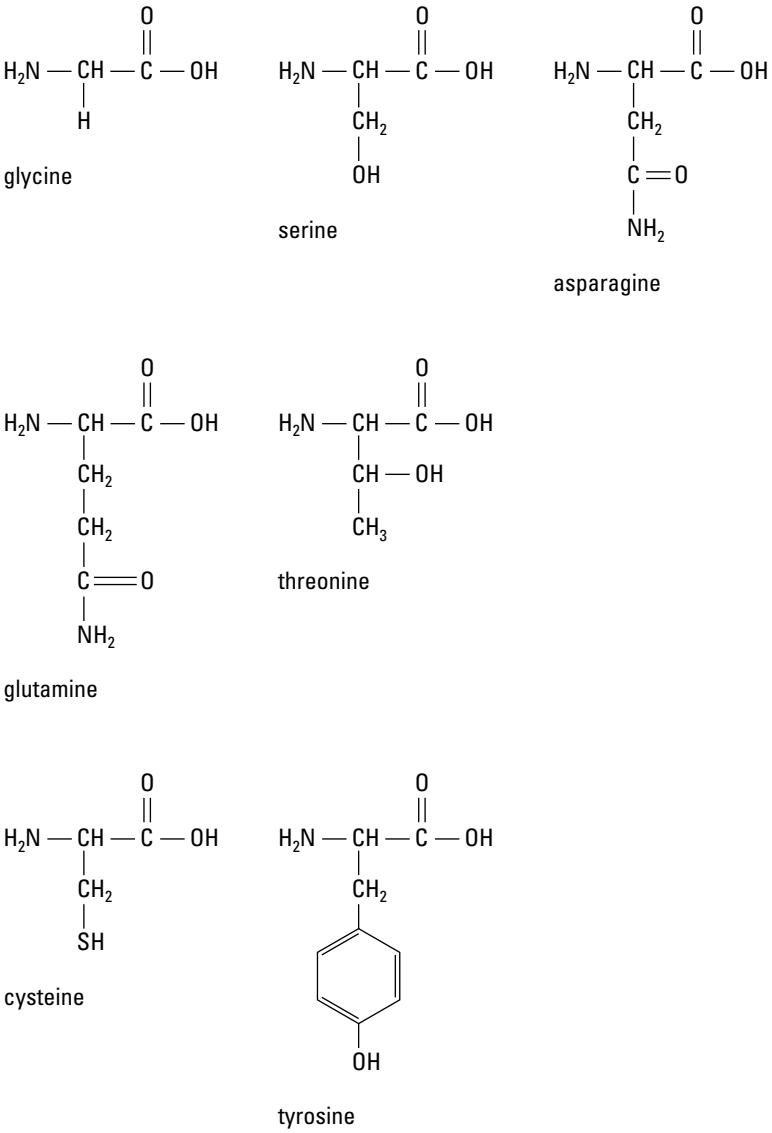


Figure 4-5: Polar and uncharged amino acids.

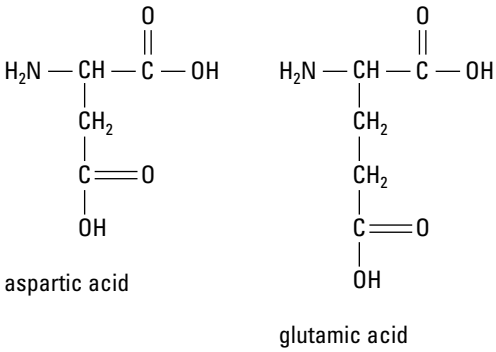


Figure 4-6: Acidic amino acids.

Basic amino acids

The basic amino acids are as follows:

- Histidine (his, H)
- Arginine (arg, R)
- Lysine (lys, K)

All of these are classified as basic amino acids, but dramatic changes in pH can affect their reactivities. This is especially true of histidine.

In all three of these amino acids, there is a basic group capable of accepting a hydrogen ion. In the case of lysine, this is a simple ammonium ion. Arginine forms the guanidinium group. Histidine forms an imidazolium group. As in the case of the acidic side chains, these side chains have a pK_a value. Both arginine and lysine are usually protonated at physiological pH values. In these cases, there is a net +1 charge present. In proteins, this net charge may be part of an ionic interaction. The pK_a of the side chain of histidine is lower than other basic groups. Protonation of histidine becomes significant at much lower pH values. In many proteins, histidine is not protonated, but is important in many enzymes in hydrogen ion transfer processes. Figure 4-7 shows these basic amino acids.

Lest We Forget: Rarer Amino Acids

In a few cases, an amino acid may undergo modification once it is incorporated into a protein. Collagen and gelatin, for example — proteins present in higher vertebrates — contain hydroxylysine and hydroxyproline. These two amino acids contain an additional –OH group on the side chain.

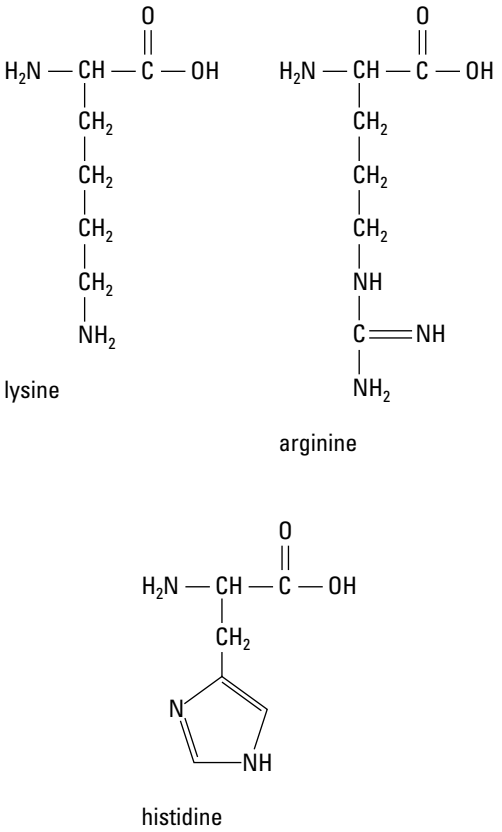


Figure 4-7: Basic amino acids.

Certain amino acids do not occur in proteins. The neurotransmitter γ-aminobutyric acid — GABA — is one example. Citrulline is the amino acid that serves as a precursor of arginine. Ornithine, homocysteine, and homoserine are important as metabolic intermediates. Figure 4-8 shows a couple of these amino acids.

Rudiments of Amino Acid Interactions

Amino acids are the ingredients used in the recipe in making a protein. Just as the individual ingredients in a recipe lead to distinct characteristics of what eventually shows up on the dinner table, the amino acids present contribute properties to proteins. And just as you cannot replace the flour in a recipe with pepper, you generally cannot replace one amino acid in a protein with another. In both cases, the final product will be different. In the next section, we show you some of the ways that amino acids interact. These interactions set the stage for our discussion of bonding among the amino acids.

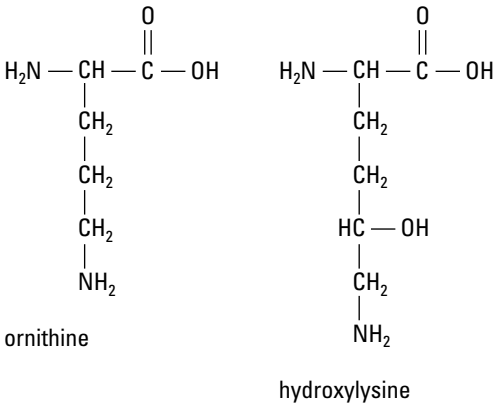


Figure 4-8: Two of the less common amino acids.

Intermolecular forces: How an amino acid reacts with other molecules

Amino acids can interact with other molecules — and we mean *any* other molecules, including fluids, other amino acids, and other biological molecules — in a variety of ways. We cover intermolecular forces in general in Chapter 3, but in this section we show you how they play out when amino acids are involved. The carboxylic acid and amine parts of the amino acids define much of the reactivity of the molecule, but the side chains can also interact with other molecules. There are three general ways in which they can interact.

- **Hydrophobic interactions:** The nonpolar side groups are hydrophobic and are attracted to each other through London dispersion forces. Nonpolar groups tend to clump together and exclude not only water but also all other types of side chains.
- **Hydrophilic reactions:** The polar and uncharged side groups are hydrophilic. The presence of a number of these groups increases the solubility of a protein. These groups hydrogen bond not only to water but also to each other. Polar groups tend to interact strongly and "push" the nonpolar groups out.
- **Ionic interactions:** The presence of acidic or basic side chains leads to ionic charges — opposite charges attract. A carboxylate group from one side chain is attracted to the ammonium ion of another side chain through an ionic interaction. This ionic bond is very strong.

The amino acid cysteine can interact with a second cysteine molecule through a different type of interaction (Figure 4-9). The mild oxidation of two cysteine sulfhydryl groups leads to the formation of cystine. A disulfide linkage joins the two amino acids with a covalent bond. Mild reduction can reverse this process.

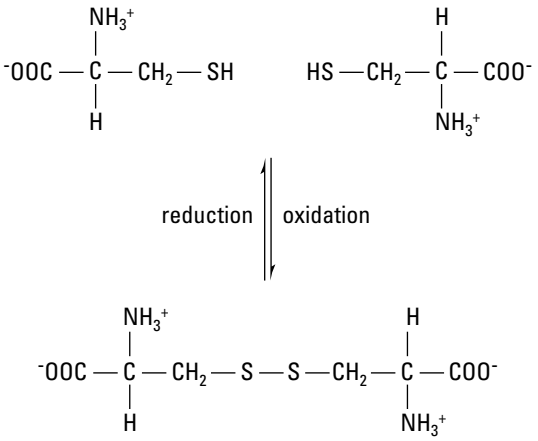


Figure 4-9: Joining two cysteine molecules to form cystine.

A hair perm utilizes an oxidation reduction reaction creating disulfide linkages. The greater the number of disulfide linkages, the curlier the hair!

Altering interactions by changing an amino acid's pH

As we discuss in Chapter 3, the function of many substances, especially biochemical ones, is dependent on pH. If you change the pH, you change some of the interactions. In this section we show how those changes affect interactions involving amino acids.

Just like any other molecule, an amino acid has two or three functional groups, depending on the amino acid. Those functional groups include those with oxygen and sulfur, those with nitrogen, and those with phosphorus. A change in pH affects one to three of those functional groups in terms of interactions. So if an amino acid has a functional group that changes from a dipole-dipole interaction to an ionic interaction.

One example of the dipole-dipole to ionic interaction change is the process of milk curdling. If you add an acid to milk, it coagulates. Casein has an isoelectric point at 4.6 pH, so that adding an acid causes the formation of ionic bonds among the molecules. This works against the dipole-dipole interactions with water, so that the protein precipitates.

The pK_a values for the various groups present in the different amino acids are shown in Table 4-1. If the pH of the solution matches one of these values, then half the species is in the protonated form and half is in the deprotonated

form. At a lower pH, more than half is protonated, whereas at a higher pH more than half is deprotonated.

The pH dependence of the protonation of amino acids aids in their separation and identification. Because the amino acids use the carboxylic acid and amine ends when they join to form a protein, only the pK_a values of the side chains are important in additional interactions and reactions.

Table 4-1 pK_a Values for the Amino Acids

Amino acid	*pK_a –COOH*	*pK_a –NH_3^+*	*pK_a R group*
Alanine	2.35	9.69	
Arginine	2.17	9.04	12.48
Asparagine	2.02	8.8	
Aspartic acid	2.09	9.82	3.86
Cysteine	1.71	10.78	8.33
Glutamic acid	2.19	9.67	4.25
Glutamine	2.17	9.13	
Glycine	2.34	9.6	
Histidine	1.82	9.17	6.0
Isoleucine	2.36	6.68	
Leucine	2.36	9.60	
Lysine	2.18	8.95	10.53
Methionine	2.28	9.21	
Phenylalanine	1.83	9.13	
Proline	1.99	10.60	
Serine	2.21	9.15	
Threonine	2.63	10.43	
Tryptophan	2.38	9.39	
Tyrosine	2.20	9.11	10.07
Valine	2.32	9.62	

Combining Amino Acids: How It Works

A *protein* is a string of at least 150 amino acids (residues) joined. We cover the fundamentals about protein creation in Chapter 5, but before you dive into that topic, this section gives you a solid understanding of how two amino acids join together in the first place, and how additional amino acids link onto the chain gang. The process is reversible (as in digestion).

When drawing the chemical structures of amino acids and their bonds, the standard convention is to first draw the structures from the ammonium group of the first amino acid (the N-terminal residue), starting at the left, and continuing the drawing to the right, ending with the carboxylate group (C-terminal residue) of the last amino acid.

The peptide bond and the dipeptide

One of the most important types of bonds in all of biochemistry is the *peptide bond.* As you will see, it is this type of bond that will be used in the synthesis of proteins. The interaction of two amino acids at the body's pH results in the formation of a peptide bond as illustrated in Figure 4-10.

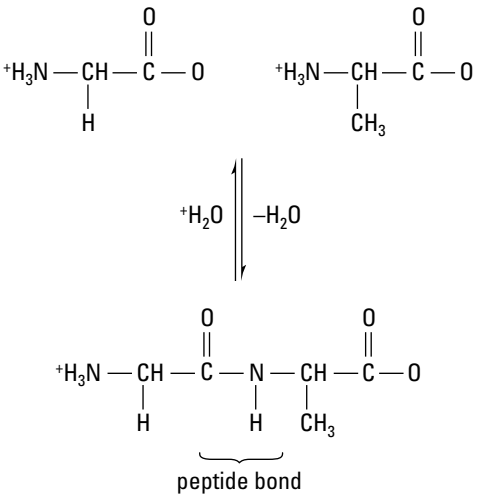


Figure 4-10: The formation of a peptide bond.

The two residues react to expel a water molecule, the same dehydration you used so much in organic chemistry. The reverse of this condensation reaction is hydrolysis. The resultant amide group is a peptide bond. The presence of two amino acid residues means the product is a dipeptide.

The peptide bond is a *flat* (planar) structure. It is stabilized by our old organic friend, *resonance.* Figure 4-11 illustrates the stabilization. The resonance increases the polarity of the nitrogen and oxygen. This increase in polarity leads to hydrogen bonds that are much stronger than most other hydrogen bonds. The double bond character between the carbon and the nitrogen restricts rotation about this bond. That's why the peptide bond is planar.

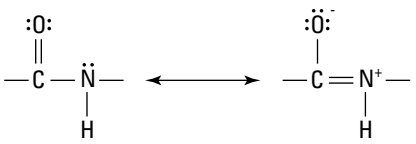


Figure 4-11: Resonance stabilization of a peptide bond.

Tripeptide: adding an amino acid to a dipeptide

A repetition of the process illustrated in Figure 4-10 joins a third amino acid to produce a *tripeptide.* For example, combining glycine, alanine, and serine yields the illustration in Figure 4-12. Notice that everything begins with the N-terminal residue and ends with the C-terminal residue. (You could designate this tripeptide as *gly-ala-ser* using the three letter abbreviations.)

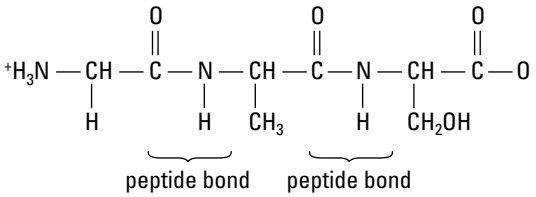


Figure 4-12: A tripeptide.

The repetition of the process of linking amino acids hundreds or thousands of times produces a protein. In the next chapter, we cover that topic in full.

Chapter 5

Protein Structure and Function

In This Chapter

- Finding out about the structure of proteins
- Understanding amino acid sequencing in proteins
- Learning about applications of protein sequencing

In Chapter 4, we show you how amino acids combine through the use of a peptide bond, and we mention there that if at least 150 or so amino acids join hands, they rise to the rank of a protein. However, distinguishing an amino acid chain as a protein isn't exactly simple — just as written English is an extremely diverse set of words made by combining letters from an alphabet of just 26 letters, proteins are an extremely diverse set of biochemicals made by combining 20 different amino acids.

In this chapter, we show you more about these proteins, including the four types of protein structure that determine a protein's function and the sequence of amino acids in a particular protein.

There are two general categories of proteins:

- **Fibrous proteins** are found only in animals. They usually serve as structural entities — for example, connective tissue, tendons, and muscle fiber. They are normally insoluble in water.
- **Globular proteins** usually do not serve a structural function — they act as transporters, like hemoglobin, and are often enzymes. They are usually water-soluble.

Proteins are utilized in living organisms in a number of ways, such as:

- **Structure:** Skin and bone contain collagen, a fibrous protein.
- **Catalysis:** These proteins, called enzymes, allow reactions to occur in the organism under mild conditions and with great specificity.
- **Movement:** Proteins make up a large protein of muscle fiber and help in the movement of various parts of our bodies.

- **Transport:** These proteins transport small molecules through the organism. Hemoglobin, the protein that transports oxygen to the cells, is a transport protein.
- **Hormones:** Proteins called hormones help regulate cell growth.
- **Protection:** Proteins called antibodies help rid the body of foreign proteins.
- **Storage:** These protein help store other substance in the organism. For example, iron is stored in the liver in a complex with the protein ferritin.
- **Regulation:** These proteins help mediate cell responses, such as the protein rhodopsin, found in the eye and involved in the vision process.

The function that a particular protein assumes is, in many cases, directly related to the structure of that protein. Proteins may have as many as four levels of structure (key word being *levels,* not different structures), each of which places the components into a position where these intermolecular forces can interact most advantageously. The levels are simply labeled primary, secondary, tertiary, and quaternary. Primary is the most fundamental level that all proteins have, and quaternary is the most specific level that only some proteins have. Intermolecular forces themselves are important to the function of a protein, of course, but the arrangement of the molecules is even more significant.

If present, the secondary, tertiary, and quaternary structures of a protein may be destroyed — in a number of ways:

- Heating (cooking) can break hydrogen bonds.
- Changing the pH can protonate or deprotonate the molecule and interrupt ionic interactions.
- Reducing agents can break disulfide linkages.

In some cases, the process may be reversible.

Primary Structure: The Structure Level All Proteins Have

The primary structure of a protein is simply the sequence of amino acids comprising the molecule. The primary structure of a protein is the amino acid sequence within the molecule. All proteins have a primary structure, because all proteins by definition consist of a sequence of amino acids. The primary structure serves as the foundation upon which all higher levels of protein structure build.

Next we take a look at how a protein is assembled from its building blocks, the amino acids.

Building a protein: Outlining the process

During the synthesis of a protein, the chain of amino acids is built one link at a time, roughly as follows:

1. **The transfer RNA (tRNA) molecule transfers specific amino acids to the mitochondria of the cell to connect to the growing chain.**
2. **Each amino acid joins to the chain through the formation of a peptide bond.** (See Chapter 4 for more on peptide bonds.)
3. **The first peptide bond joins two amino acids to form a dipeptide.**
4. **The second peptide bond joins three amino acids to produce a tripeptide.**
5. **This process continues hundreds, if not thousands, of times to produce a polypeptide — a protein.**

When two or more amino acids combine, a molecule of water is removed. What remains of each amino acid is called a *residue.* They lack a hydrogen atom on the amino group, or an –OH on the carboxyl group, or both.

The cell's DNA ultimately controls the sequence of amino acids. This information goes from the DNA to the messenger RNA (mRNA), which serves as the template for the creation of the primary structure of the protein. It is necessary to supply energy, as we will see later, to synthesize the protein.

Organizing the amino acids

One end of the primary structure has an amino group, and the other end has a carboxylate group. By convention, the end with the amino group is considered the "beginning" of the protein. Drawing, naming, numbering, and other treatments of the primary structure always begin with the amino end (called the *N-terminal*) and stop with the carboxylate end (the *C-terminal*). For example, in the hexapeptide Met-Thr-Ser-Val-Asp-Lys (see Chapter 4 for a list of the amino acids and their abbreviations), methionine (Met) is the N-terminal amino acid, and lysine (Lys) is the C-terminal amino acid. Note that reversing the sequence to Lys-Asp-Val-Ser-Thr-Met also gives a hexapeptide with the same composition but with different chemical properties because you initially started with a different amino acid. Therefore an amino acid that lost a hydrogen in one sequence will lose an –OH in the other.

The polypeptide chain has a backbone consisting of the same, rather simple, repeating unit. Variations take place in the form of side chains — the R groups of the amino acids. You can see this repeating sequence in Figure 5-1. Notice that the repeating unit (indicated by the brackets) is the amino-carbon-carbonyl sequence and that there can be different R groups attached to the carbon unit of this backbone.

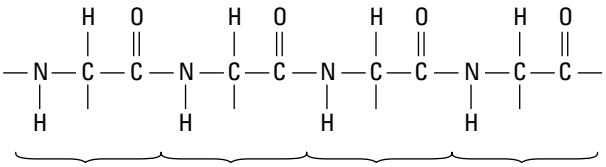


Figure 5-1: Repeating sequence of the protein backbone.

The protein backbone has many places where hydrogen bonds may form. Every residue — other than the amino acid proline — has an NH, which may serve as a *donor* to a hydrogen bond. And every residue has a carbonyl group, which can serve as the *acceptor* of a hydrogen bond. The presence of donors and acceptors leads to the possibility of forming numerous hydrogen bonds.

Each of the peptide bonds exhibits no free rotation about the carbon-nitrogen bond because of the contribution of the resonance form, which has a double bond. Thus, there is a planar unit of four atoms, and in almost all cases, the oxygen atom is trans to the hydrogen atom. The remainder of the backbone can rotate. The ability to rotate or not influences how the three-dimensional structure of the protein is established. There are restrictions to this rotation because the side-chains can "bump" into each other — called *steric* hindrance. The rigidity of the peptide bond and rotation restrictions lower the entropy of the three-dimensional structure of a protein relative to a random chain of amino acids. Lowering the entropy helps stabilize the structure.

Example: The primary structure of insulin

The first determination of the primary structure of a protein was that of bovine insulin, the structure of which appears in Figure 5-2. Since this landmark determination, the primary structures of more than 100,000 proteins have been determined. In all cases, the protein has a unique primary structure.

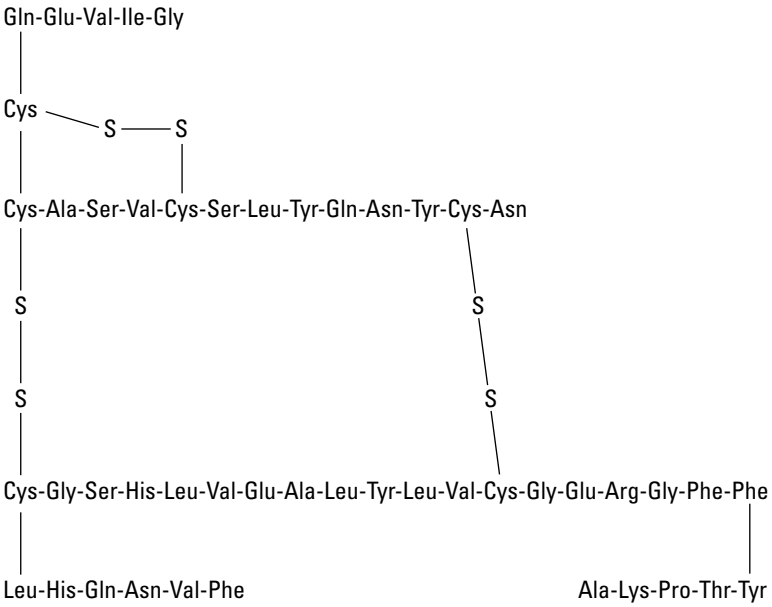


Figure 5-2: Structure of bovine insulin.

Secondary Structure: A Structure Level Most Proteins Have

It is possible for one peptide bond to form a hydrogen bond to another peptide bond. In general, the formation of these hydrogen bonds leads to the secondary structure of a protein. The secondary structure is the result of many hydrogen bonds, not just one. The hydrogen bonds are intramolecular, that is between segments of the same molecule, as shown in Figure 5–3:

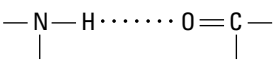


Figure 5-3: Hydrogen bonding between two peptide bonds.

The α-helix and β-pleated sheet are the secondary structures that result from this hydrogen bonding. Secondary structures may be only a small portion of the structure of a protein or can make up 75 percent or more.

The α-helix

In the α-helix, the primary structure twists into a tightly wound, spring or rod-like structure. Each turn consists of 3.6 amino acid residues. These turns allow hydrogen bonding between residues spaced four apart. Every peptide bond participates in two hydrogen bonds: one from an NH to a neighboring carbonyl, and one from a neighboring NH to the carbonyl (Figure 5-4).

Structurally, the helices may be either right-handed or left-handed (see Chapter 3 for more on handedness). Essentially all known polypeptides are right-handed. Slightly more steric hindrance is present in a left-handed helix, and the additional steric hindrance makes the structure less stable. Keratin — the protein of fur, hair, and nails — consists of three right-handed α-helices wrapped around each other in a left-handed coil.

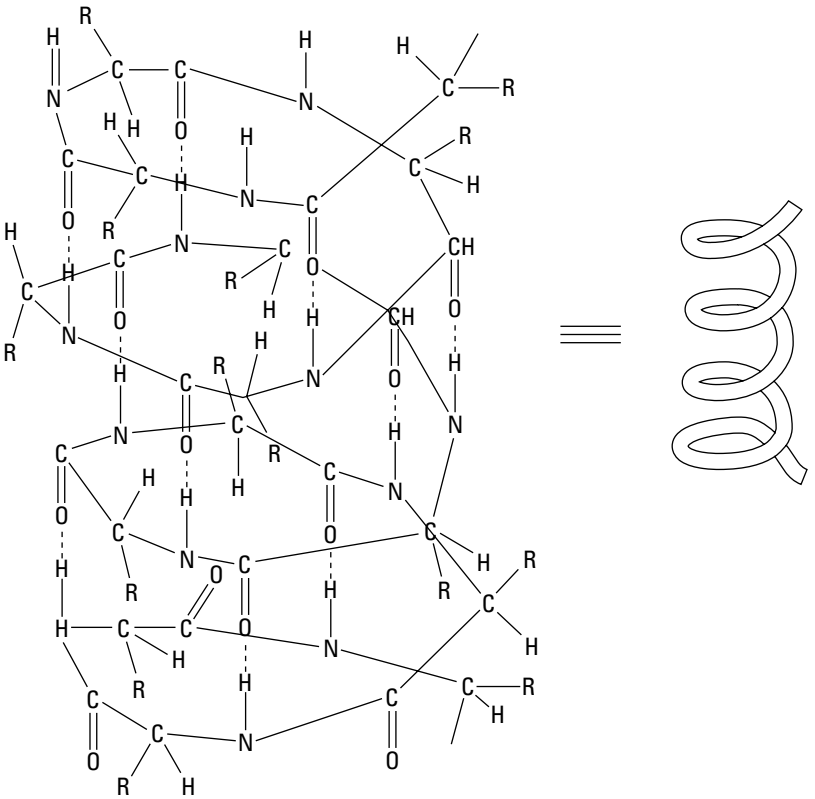


Figure 5-4: The α-helix.

Certain amino acids destabilize the α-helix. Proline, for example, creates bends or "kinks" in the primary structure, which inhibit the formation of a regular pattern of hydrogen bonds. A group of isoleucine residues disrupts the secondary structure because of the steric hindrance caused by their bulky R groups. The small R group of glycine, only an H, allows too much freedom of movement, which leads to a destabilization of the helix. A concentration of aspartic acid and/or glutamic acid residues also destabilizes the structure because the negative charges on the side chains repel each other. Other residues that destabilize the helix, for similar reasons, are lysine, arginine, serine, and threonine.

The β-pleated sheet

The β-pleated sheet, or simply the β sheet, is the other major secondary protein structure. Here, the primary structure is extended instead of tightly winding into a helix. There are two forms of this structure, known as the *parallel* β-pleated sheet and the *anti-parallel* β-pleated sheet. Again, hydrogen bonds are the source of these structures. A β-pleated sheet forms when two or more strands link by hydrogen bonds. The strands are different parts of the same primary structure.

In the parallel structure, the adjacent polypeptide strands align along the same direction from N-terminal end to C-terminal end. In the anti-parallel structure, the alignment is such that one strand goes from N-terminal end to C-terminal end, while the adjacent strand goes from C-terminal end to N-terminal end (Figure 5-5).

In the β-pleated sheet structures, the side chains of adjacent amino acids point in opposite directions. The hydrogen bonding pattern in the parallel structure is the more complicated. Here, the NH group of one residue links to a CO on the adjacent strand, whereas the CO of the first residue links to the NH on the adjacent strand that is two residues down the strand. In the anti-parallel structure, the NH and CO groups of one residue link to the respective CO and NH groups of one residue on the adjacent strand.

Schematically, broad arrows indicate the presence of β-pleated sheets. If the arrows point in the same direction, it is the parallel structure, and if they point in opposite directions, it is the anti-parallel structure. The sheets are typically 4 or 5 strands wide, but 10 or more strands are possible. The arrangements may be purely parallel, purely anti-parallel, or mixed (refer to Figure 5-5).

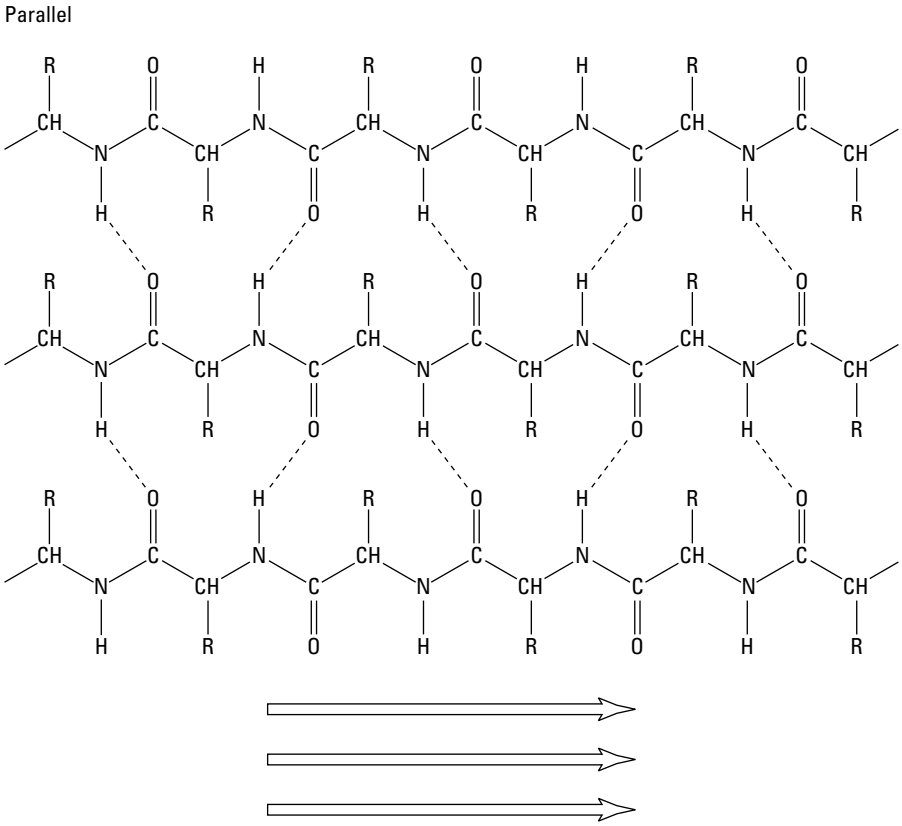


Figure 5-5: Parallel and anti-parallel β-pleated sheet structures.

β-turns and the Ω-loops

There are additional secondary structures involving hydrogen bonding between peptide bonds; these are much smaller units. The best known are the β-turn — or *hairpin bend* — and the Ω-loop. The hairpin bend is simply a bend in the primary structure held in place by a hydrogen bond. The Ω-loop gets its name because of the loose similarity of its shape to the Greek letter. Both are found on the exterior of proteins.

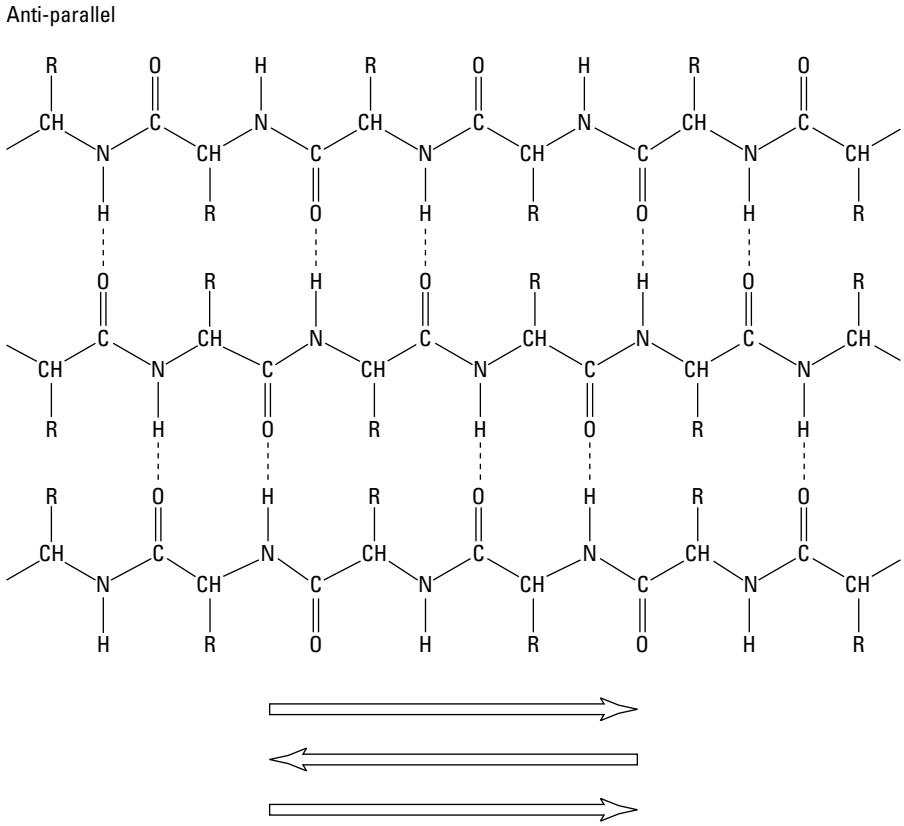


Figure 5-5: (continued)

Tertiary Structure: A Structure Level Many Proteins Have

The overall shape of a protein is determined by its primary and secondary structures along with interactions between the side chains. This gives rise to what is called the protein's *tertiary structure.* Nonpolar side chains are hydrophobic and, although repelled by water, are attracted to each other. Polar side chains attract other polar side chains through either dipole-dipole forces or hydrogen bonds.

For example, both aspartic acid and glutamic acid yield side chains with a negative charge that are strongly attracted to the positive charges in the side chains of lysine and arginine. Two cysteine residues can connect by forming a disulfide linkage — a covalent bond (Figure 5-6).

What induces a protein to adopt a very specific tertiary structure? Examination of the structures of many proteins shows a preponderance of nonpolar side chains in the interior with a large number of polar or ionic side chains on the exterior. In an aqueous environment, the hydrophobic (nonpolar) groups induce the protein to fold upon itself, burying the hydrophobic groups away from the water and leaving the hydrophilic groups adjacent to water. The result is similar in structure to a micelle.

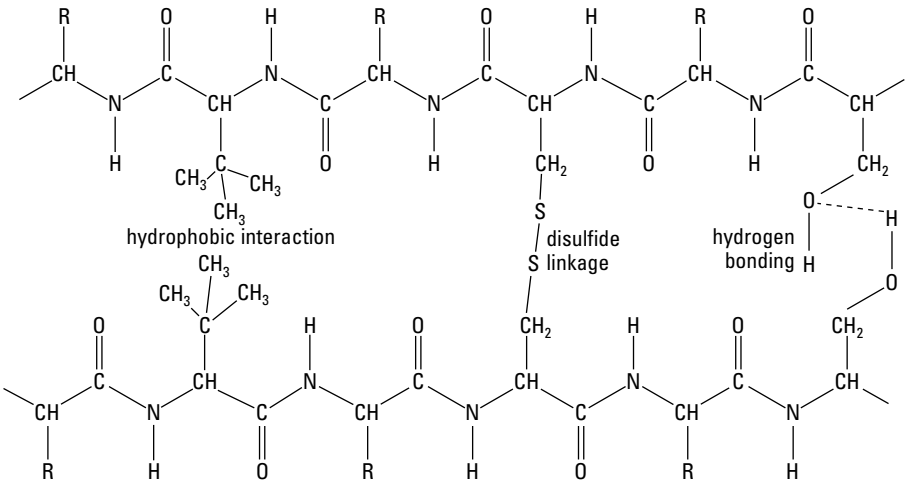


Figure 5-6: Some tertiary structures appearing in proteins.

Quaternary Structure: A Structure Level Some Proteins Have

The quaternary structure found in some proteins results from interactions between two or more polypeptide chains — interactions that are usually the same as those that give rise to the tertiary structure. These interactions include hydrogen bonding and disulfide bonds. This quaternary structure locks the complex of proteins into a specific geometry. An example is hemoglobin, which has four polypeptide chains. There are two identical α-chains and two identical β-chains. (The designations α and β simply refer to two different proteins and not to secondary structures.)

Dissecting a Protein for Study

The previous sections have discussed the different types of protein structure. Now it is time to see how a biochemist goes about determining the structure(s) of a particular protein.

Additional information about the structure of a protein comes from immunology. An animal generates an antibody in response to a foreign substance known as an *antigen*. *Antibodies* are proteins found in the blood serum. Exposure to diseases, certain chemicals, and allergies induce the formation of specific antibodies. These antigens collect on the surface of red blood cells. Every antigen has a specific antibody.

Antibodies are very specific and have a strong affinity for their specific antigens, recognizing specific amino acid sequences on the antigens. Animals have a large number of antibodies present in their bodies, based on their environmental history. One application of antibodies and antigens is in the analysis of bloods, specifically in the field of forensics investigations (see nearby "Forensics: Analysis of bloodstains" sidebar).

Separating proteins within a cell and purifying them

There are thousands of different proteins in each cell. In order to examine and study one of them, you need to separate it from all the others. The methods of separating proteins are, in general, applicable to all other types of biochemicals. Initially, simple filtration and solubility can remove gross impurities, but much more needs to be done before the sample is pure. The key separation and purification methods depend on two physical properties of the proteins: size and charge.

Separating proteins by size

Methods relying on separation by protein size and mass include ultrafiltration, ultracentrifugation, and size exclusion chromatography. *Ultrafiltration* is a modification of dialysis in which molecules smaller than a certain size diffuse through a semipermeable membrane, and larger ones don't. Ultrafiltration can separate smaller molecules from larger impurities or larger molecules from smaller impurities.

In *ultracentrifugation,* a powerful centrifuge causes heavier molecules to sink faster and, which allows their separation — much as the lighter water is separated from the heavier lettuce in a salad spinner. Ultracentrifugation also gives the molar mass of the protein.

In *size exclusion chromatography,* also known as *molecular sieve chromatography* or *gel filtration chromatography,* a solution passes through a chromatography column filled with porous beads. Molecules that are too large for the pores pass straight through. Molecules that may enter the pores are slowed. The molecules that may enter the pores undergo separation depending on how easily they can enter.

Forensics: Analysis of bloodstains

The study of proteins has many applications to forensics. One of them is the examination of bloodstains, blood being the most common form of evidence examined by a forensic serologist. The presence of blood can link a suspect to both a victim and a crime scene. Bloodstain patterns can also give evidence of how a violent attack took place. Criminals recognize the significance of this evidence and often try to conceal it.

Blood is mostly water, but it also contains a number of additional materials including cells, proteins, and enzymes. The fluid portion, or *plasma,* is mostly water. The *serum* is yellowish and contains platelets and white blood cells. The *platelets,* or red blood cells, outnumber the white blood cells by about 500 to 1. White blood cells are medically important, whereas red blood cells and, to a lesser extent, serum are important to the forensic serologist. Because blood quickly clots when exposed to air, serologists must separate the serum from the clotted material. The serum contains antibodies that have forensic applications, and red blood cells have substances such as antigens on their surfaces that also have forensic applications. Antibodies and antigens are the keys to forensic serology: Even identical twins with identical DNA have different antibodies. As you know from this chapter, antibodies, and some antigens, are proteins, and this is why methods of studying proteins are important to their analysis.

Analysis of bloodstains initially attempts to answer five questions.

- **Is this a blood sample?** To answer this question, the investigator can use a number of tests. The generic term for a test of this type is a *presumptive* test. The Kastle-Meyer test uses phenolphthalein, which, when it comes into contact with hemoglobin, or a few other substances, forms a bright pink color from the release of peroxidase enzymes. The luminol test is useful in detecting invisible bloodstains because, in contact with blood, or a few other chemicals, luminol emits light, which can be seen in a darkened room. The Wagenhaar, Takayama, and Teichman tests take advantage of the fact that long-dried blood will crystallize or can be induced to crystallize.
- **Is the blood from a human or an animal?** The forensic investigator answers this question (and the next one, if applicable) by means of an *antiserum* test. It is important to know whether the blood came from a human or an animal such as a pet. The standard test is the precipitin test. Injecting human blood into an animal results in the production of antibodies in the animal's bloodstream, and isolating these antibodies from the animal's blood yields an antiserum. If human antiserum creates clotting in a blood sample, the sample must be human.
- **If the sample is from an animal, what is the species?** It is possible to create animal antiserums in an analogous manner, and test for each type of animal.
- **If the blood is from a human, what is the blood type?** The procedure for answering this depends on the quantity and quality of the sample. If the quality is good, *direct* typing is done — otherwise, *indirect* typing is used. (Direct typing, to classify blood in the A-B-O system, is discussed in this chapter's other sidebar.) A dried bloodstain normally requires indirect typing. The most common indirect typing method is the absorption-elution test. Treatment of a sample with antiserum antibodies gives a solution which, upon addition to a known sample, causes coagulation.

- **Is it possible to determine the sex, race, and age of the source of the blood?** Here the answers become less precise. Clotting and crystallization indicate age. Testing for testosterone levels and chromosome testing can determine sex. And certain controversial, racial genetic markers based on protein and enzyme tests may indicate race.

  Other body fluids may contain the same antibodies and antigens found in blood. Therefore, similar tests work on these fluids as well.

Separating proteins by charge

Methods of separating proteins relying on the charge of the protein include solubility, ion exchange chromatography, and electrophoresis. Each of these methods is pH dependent.

Proteins are least soluble at their isoelectric point. (The *isoelectric point* is the pH where the net charge on the protein is 0.) At the isoelectric point, many proteins precipitate from solution. At a pH below the isoelectric point, the protein has a net positive charge, whereas a pH above the isoelectric point imparts a net negative charge. The magnitude of the charge depends on the pH and the identity of the protein. Therefore, two proteins coincidently having the same isoelectric point will not necessarily have the same net charge at a pH that is one unit lower than the isoelectric point.

Both ion exchange chromatography and electrophoresis take advantage of the net charge. In *ion exchange chromatography,* the greater the magnitude of the charge, the slower a protein moves through a column — this is similar to the ion-exchange process that occurs in water-softening units.

In *electrophoresis,* the sample solution is placed in an electrostatic field. Molecules with no net charge do not move, but species with a net positive charge move toward the negative end, and those with a net negative charge move toward the positive end. The magnitude of the net charge determines how fast the species moves. Other factors influence the rate of movement, but the charge is the key. There are numerous modifications of electrophoresis.

In protein analysis, rarely do biochemists use only one single technique. They commonly use several in order to confirm their findings.

Digging into the details: Uncovering a protein's amino acid sequence

Once a pure sample of protein is available, it is possible to begin determining its amino acid sequence, in order to identify the specific protein. The general

procedure for doing so, with slight modification, works for other biochemicals as well:

Step 1: Separating and purifying the polypeptide chains

If you determine that more than one polypeptide chain is present in the protein, you need to separate and purify the chains so you can sequence them individually. (Because many proteins only have one polypeptide chain, this step is not always necessary.) Denaturing the protein, disrupting its three-dimensional structure without breaking the peptide bonds, using pH extremes will normally suffice. If disulfide linkages are present between the chains, apply the procedure outlined in Step 2 to separate the chains for isolation.

Step 2: Slashing intrachain disulfide linkages

Step 2 requires breaking (cleaving) the disulfide linkages. A simple reduction accomplishes this. However, the linkages may reform later, so it is necessary to cleave the linkages and prevent their reformation via reductive cleavage followed by alkylation. Oxidative cleavage, where oxidation of the sulfur to $-SO_3^-$ occurs, also prevents a reversal of the process.

Step 3: Determining amino acid concentration of the chain

Step 3 is easily accomplished using an *amino acid analyzer*, an automated instrument that can determine the amino acid composition of a protein in less than an hour. The instrument requires less than a nanomole of protein. The analyzer's output is the percentages of each of the amino acids present.

Step 4: Identifying the terminal amino acids

Step 4 not only identifies the terminal amino acids but also indicates whether more than one chain is present. A polypeptide chain only has one N-terminal and one C-terminal amino acid. Therefore, if more than one N- or C-terminal amino acid is present, there must be more than one polypeptide chain.

It is possible to identify the N-terminal residue in a number of ways. In general, procedures begin by adding a reagent that reacts with the N-terminal amino acid and tags it. Subsequent hydrolysis destroys the polypeptide, allowing separation of the tagged residue and its identification. Such methods use Sanger's reagent, dansyl chloride, and leucine aminopeptidase. The method of choice nowadays is called the Edman degradation. This method, as do other methods, tags the N-terminal residue; however, only the terminal amino acid is cleaved from the chain, so the remainder of the chain is not destroyed as in other methods. It is possible to repeat the procedure on the shortened chain to determine the next residue. In principle, repetition of the Edman degradation can yield the entire sequence, but, in most cases, determination of the first 30 to 60 residues is the limit.

It is also possible to determine the C-terminal residue by tagging. The akabori reaction (hydrazinolysis) and reduction with lithium aluminum hydride tag the C-terminal residue. It is also possible to selectively cleave the C-terminal residue using the enzyme carboxypeptidase, a variety of which are available. Unfortunately, the enzyme doesn't stop with one cleavage — given sufficient time, it proceeds down the entire polypeptide chain.

Steps 5 and 6: Breaking the chain into smaller pieces

In Step 5, you cleave the polypeptide into smaller fragments and determine the amino acid composition and sequence of each fragment. Step 6 repeats Step 5 using a different cleavage procedure to give a different set of fragments. Steps 5 and 6 break the chain into smaller pieces to ease identification.

Most of the methods here employ enzymes; however, other less-specific methods are useful in some cases. Partial acid hydrolysis randomly cleaves the protein chain into a number of fragments. Trypsin, a digestive enzyme, specifically cleaves on the C-side of arginine or lysine. Using trypsin gives additional information that the total number of arginine and lysine residues present is one less than the number of fragments generated. The digestive enzyme chymotrypsin preferentially cleaves residues containing aromatic rings (tyrosine, phenylalanine, and tryptophan). It slowly cleaves other residues especially leucine. Clostripain cleaves positively charged amino acids, especially arginine. It cleaves lysine more slowly. Fragments with a C-terminal aspartic acid or glutamic acid form from the interaction of staphylococcal protease on a protein in a phosphate buffer. In the presence of bicarbonate or acetate buffer, only C-terminal glutamic acid fragments result. A number of less specific enzymes can complete the breakdown of the fragments, including elastase, subtilisin, thermolysin, pepsin, and papain.

Chemical methods of breaking up the fragments include treatment with cyanogen bromide, hydroxylamine, and heating an acidic solution. Cyanogen bromide specifically attacks methionine. Hydroxylamine specifically attacks asparagine-glycine bonds. If a solution at pH = 2.5 is heated to 40°C, selective cleavage of aspartic acid-proline bonds occurs.

It is possible to apply the Edman degradation on each of the fragments. This can simplify the determination of the sequence of a large protein.

Step 7: Combining information to get the total sequence

Step 7 is where the information from the various procedures comes together. For example, look at a simple octapeptide fragment from a protein. This fragment gave, upon complete hydrolysis, one molecule each of alanine (Ala), aspartic acid (Asp), glycine (Gly), lysine (Lys), phenylalanine (Phe), and valine (Val), and two molecules of cysteine (Cys). The following fragments were

Basics of blood typing

The determination of blood type in the A-B-O system, first begun in 1901, is based on antigen-antibody reactions. Over the years, additional reactions have been discovered. More than 256 antigens are known, leading to 23 different blood groups. Each blood group is defined by the antibodies present in the serum and the antigens present on the red blood cells.

In basic blood typing, one needs two antiserums, labeled *anti-A* and *anti-B*. Adding a drop of one of these to a blood sample causes coagulation if the appropriate antigens are present. Anti-A interacts with both A and AB blood. Anti-B interacts with both B and AB blood. Neither interacts with type O blood. The approximate distribution of the different blood types is: 43–45 percent type O; 40–42 percent type A; 10–12 percent B; and 3–5 percent AB. Subgrouping is also possible with designations such as O1 and O2. There are other very rare types as well.

The Rh factor provides an additional means of subdividing blood. The *Rh factor* (the name comes from the rhesus monkey) is an antigen on the surface of red blood cells. A person with a positive Rh factor contains a protein (antibody) that is also present in the bloodstream of the rhesus monkey. About 85 percent humans are Rh positive. A person lacking this protein is, naturally, Rh negative. Assigning a blood sample as Rh positive or Rh negative is a useful simplification. There are about 30 possible combinations of factors.

Additional factors can determine whether blood belongs to a specific individual: the identification of other proteins and enzymes present in the blood. A forensic serologist (see this chapter's other sidebar for more) does this level of testing in every case where the quality of the sample allows. One of the characteristics of proteins or enzymes in the blood is *polymorphism*, or the ability to be present as isoenzymes. Polymorphism means that the protein may exist in different forms or variants. One well-known example is the polymorphism of hemoglobin into the form causing sickle cell anemia. Some well-recognized polymorphisms are:

Adenyl kinase	AK
Adenosine deaminase	ADA
Erythrocyte acid phosphatase	EAP
Esterase D	EsD
Glucose-6-phosphate dehydrogenase	G-6-PD
Glutamic pyruvate transaminase	GPT
Phosphoglucomutase	PGM 2-1
6-phosphogluconate dehydrogenase	6-PGD
Transferrin	Tf

The distribution of each of these *polymorphs* in the population is well established. The determination of each of these additional factors narrows down the number of possible individuals.

isolated after partial hydrolysis: Gly-Cys, Phe-Val-Gly, Cys-Asp, Cys-Ala, Lys-Cys, and Cys-Asp-Lys. Now we match the fragments, deduce the amino acid sequence in the octapeptide, and write a primary structure for the peptide:

Cys-Asp Lys-Cys

Cys-Asp-Lys Cys-Ala

Gly-Cys

Phe-Val-Gly

Phe-Val-Gly-Cys-Asp-Lys-Cys-Ala

Step 8: Locating the disulfide linkages

Step 8 does not specifically deal with the primary structure of the protein, but it is related. If the disulfide linkages are left intact by skipping Step 2, different fragments result. x-ray diffraction analysis can locate each amino acid residue. This can be used to determine the overall shape of a protein. In some cases, more detailed structural information can be determined by sophisticated instrumental analysis techniques.

Chapter 6

Enzymes Kinetics: Getting There Faster

In This Chapter

- Understanding enzymes classification
- Examining kinetics
- Studying the Michaelis-Menten equation
- Comprehending enzyme inhibition and regulation

Enzymes are complex biological molecules, primarily or entirely protein, which behave as biological catalysts. As *catalysts,* they alter the rate of a chemical reaction without themselves being consumed in the reaction. Enzymes are normally very specific in their action, often targeting only one specific reacting species, known as the *substrate.*

This specificity includes *stereospecificity,* the arrangement of the substrate atoms in three-dimensional space. Stereospecificity is illustrated by the fact that if the D-glucose in your diet were replaced by its enantiomer, L-glucose, you would not be able to metabolize this otherwise identical enantiomer.

Enzymes occur in many forms. Some enzymes consist entirely of proteins, whereas others have non-protein portions known as *cofactors.* The cofactor may be a metal ion, such as magnesium, or an organic substance. We call an organic cofactor a *coenzyme* (there is no specific term for a metallic cofactor). An enzyme lacking its cofactor is an *apoenzyme,* and the combination of an apoenzyme and its cofactor is a *holoenzyme.* A metalloenzyme contains an apoenzyme and a metal ion cofactor. A tightly bound coenzyme is a prosthetic group. (Wow! We know that this is a lot of terminology, but hang in there. The key is the enzyme.)

One region on the enzyme, the *active site,* is directly responsible for interacting with the reacting molecule(s). When a reacting molecule, the substrate, binds to this active site, a reaction may occur. Other materials besides the enzyme and substrate, may be necessary for the reaction to occur.

In many cases, the cell initially produces the enzyme in an inactive form called a *proenzyme* or *zymogen,* which must undergo activation for it to function. The enzyme trypsin illustrates why it is sometimes necessary to generate an inactive form of an enzyme. Trypsin is one of the enzymes present in the stomach that is responsible for the digestion of proteins. Its production, as an inactive form, occurs in the cells of the stomach walls, and activation occurs after its release into the stomach. If trypsin were produced in the active form, it would immediately proceed to begin digesting the cell that produced it. Eating yourself is not a good thing.

The activation of the inactive form of an enzyme serves as one form of enzyme control. Inhibition is another method of enzyme control. The two general types of inhibition are competitive inhibition and noncompetitive inhibition. In *competitive* inhibition, another species competes with the substrate to interact with the active site on the enzyme. In *noncompetitive* inhibition, the other species binds to some site other than the active site. This binding alters the overall structure of the enzyme so that it no longer functions as a catalyst.

Enzyme Classification: The Best Catalyst for the Job

Ever wonder who gets to name chemicals? Well, the answer varies, but for enzymes it's the Enzyme Commission of the International Union of Biochemistry that's responsible. Common names for enzymes begin with some description of its action plus an *-ase* suffix. (Enzymes that were named before the implementation of the -ase system, such as trypsin, do not follow this convention.) The Enzyme Commission has also developed a numerical system for classifying enzymes. The names begin with EC, for Enzyme Commission, and end with four numbers, separated by decimal points, describing the enzyme. An example of this nomenclature is EC 2.7.4.4.

The first number in the EC name refers to the major enzyme *class*, and there are six major enzyme classes, summarized in Table 6-1. To continue with our example, the 2 in EC 2.7.4.4 designates the enzyme as a transferase. The second number, the 7, indicates what *group* the enzyme transfers. The third number, the first 4, indicates the *destination* of the transferred group. And the last number, the second 4, refines the information given by the third number.

Table 6-1	Six Basic Types of Enzymes
Class of Enzymes	***What They Catalyze***
Oxidoreductases	Redox reactions
Transferases	The transfer groups of atoms
Hydrolases	Hydrolysis
Lyases	Additions to a double bond, or the formation of a double bond
Isomerases	The isomerization of molecules
Ligases or synthetases	The joining of two molecules

Up one, down one: Oxidoreductases

Oxidoreductases catalyze a simultaneous oxidation and a reduction. An *oxidation* involves the increase in the oxidation state of an element, whereas a *reduction* involves the decrease in the oxidation state of an element. It is impossible to have one without the other. Examples of the types of reactions that qualify as oxidation and reduction reactions are in Table 6-2. In general, the substrate undergoes either oxidation or reduction, while the enzyme temporarily does the opposite but eventually returns to its original form.

Table 6-2	Some Possible Types of Oxidation and Reduction Reactions
Oxidation	***Reduction***
Loss of one or more electrons	Gain of one or more electrons
Addition of oxygen	Loss of oxygen
Loss of hydrogen	Gain of hydrogen

An example: Succinate dehydrogenase catalyzes the oxidation of the succinate ion. In this case, the oxidation involves the loss of two hydrogen atoms with the formation of a trans double bond. The enzyme alcohol dehydrogenase

removes two hydrogen atoms from an alcohol to produce an aldehyde. The general form, unbalanced, of these reactions is as follows:

$$^{-}OOCCH_2CH_2COO^{-} \xrightarrow{\text{succinate dehydrogenase}} {}^{-}OOCCH{=}CHCOO^{-}$$

$$CH_3CH_2OH \xrightarrow{\text{alcohol dehydrogenase}} CH_3CHO$$

You don't belong here: Transferases

The purpose of a transferase is to catalyze the transfer of a group from one molecule to another. Aminotransferase transfers an amino group, and phosphotransferase transfers a phosphoryl group. The general form, unbalanced, of these reactions appears in Figure 6-1.

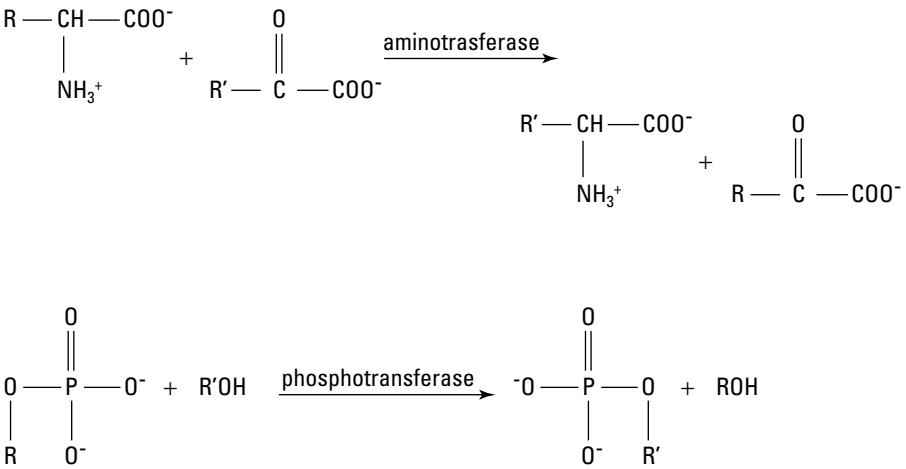
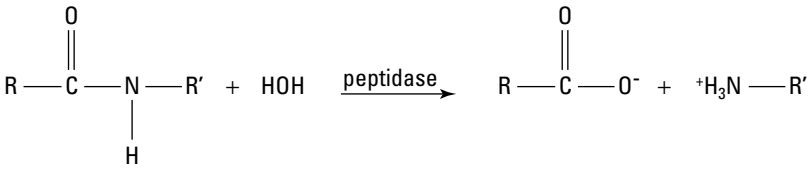


Figure 6-1: General form, unbalanced, of two transferase catalyzed reactions.

Water does it again: Hydrolases

Hydrolases catalyze the cleavage of a bond through the insertion of a water molecule (as an H and an OH). There may be a pH dependence, which results in the subsequent loss of a hydrogen ion. A *phosphatase* catalyzes the hydrolysis of a monophosphate ester, and a *peptidase* catalyzes the hydrolysis of a peptide bond. The general form of these reactions appears in Figure 6-2.



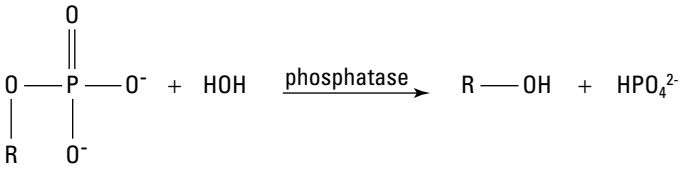
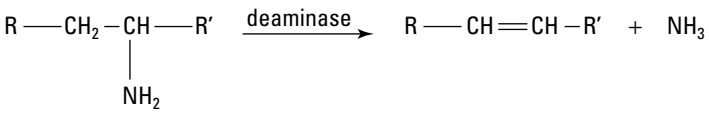


Figure 6-2: General form of two hydrolase catalyzed reactions.

Taking it apart: Lyases

Lyases catalyze the removal of a group. This process is accompanied by the formation of a double bond or the addition of a group to a double bond. A *deaminase* aids in the removal of ammonia, and a *decarboxylase* catalyzes the loss of CO_2. The general form of these reactions appears in Figure 6-3.



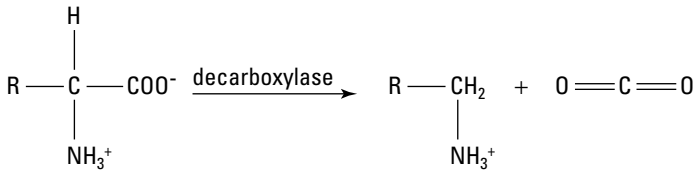


Figure 6-3: General form of two lyase catalyzed reactions.

Shuffling the deck: Isomerases

Racemase and epimerase are isomerases. *Isomerase* enzymes catalyze the conversion of one isomer to another. The *racemase* illustrated at the top of Figure 6-4 catalyzes the racemization of enantiomers. An *epimerase,* like the one at the bottom of Figure 6-4, catalizes the change of one epimer to another. Like all catalyzed reactions, these are equilibrium processes.

Putting it together: Ligases

Ligase enzymes catalyze reactions leading to the joining of two molecules in which a covalent bond forms between the two molecules. The process often utilizes high-energy bonds such as in ATP. Figure 6-5 illustrates the action of two ligases, pyruvate carboxylase and acetyl-CoA synthetase. *Pyruvate carboxylase* catalyzes the formation of a C-C bond. *Acetyl-CoA synthetase* catalyzes the formation of a C-S bond.

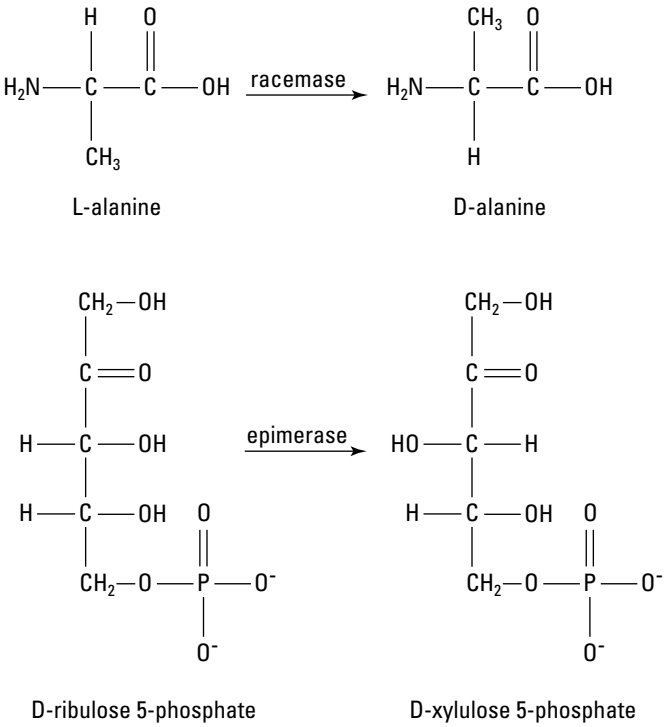


Figure 6-4: Examples of isomerase reactions catalyzed by a racemase and an epimerase.

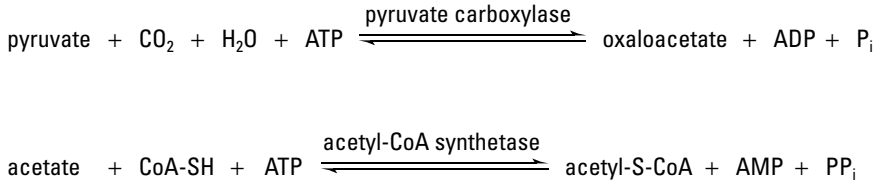


Figure 6-5: Reactions illustrating the action of the ligases pyruvate carboxylase and acetyl-CoA synthetase.

Enzymes as Catalysts: When Fast Is Not Fast Enough

The action of an enzyme begins with the formation of an enzyme-substrate complex. In this formation, the substrate in some way binds to the active site of the enzyme. The interaction between the enzyme and the substrate must, in some way, facilitate the reaction, and it opens a new reaction pathway.

The active site is typically a very small part of the overall enzyme structure. The amino acid residues comprising the active site may come from widely separated regions of the protein (primary structure), and it is only through interactions leading to higher structure levels that they are brought close together. Amino acid residues not in the active site serve many different functions that aid the function of the enzyme.

Models of catalysis: Lock and key versus induced-fit

The first attempt at explaining this process led to the *Lock and Key Model,* in which the substrate behaves as a key that fits into a lock, the enzyme (Figure 6-6). The Lock and Key Model, to a certain degree, explains the specificity of enzymes. Just as only the right key will fit into a lock, only the right substrate fits into the enzyme.

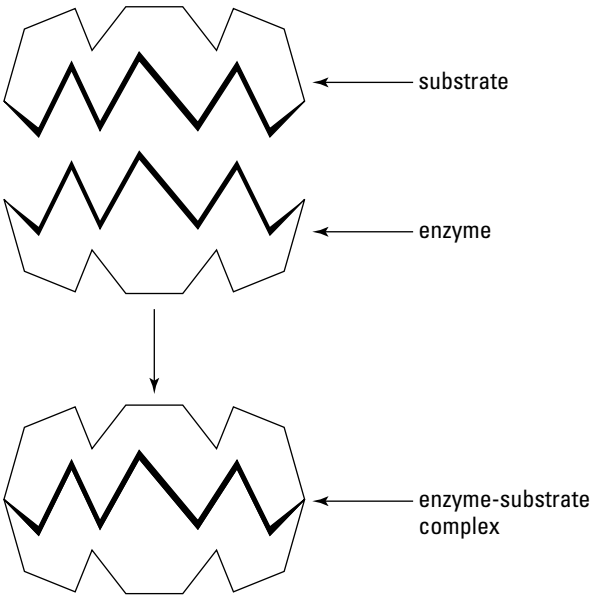


Figure 6-6: The Lock and Key Model of enzyme catalysis.

One limitation of the Lock and Key Model is that it does not explain why the reaction actually occurs, and another is that enzymes are flexible and not rigid as this theory implies.

The *Induced-Fit Model* overcomes some of the limitations of the Lock and Key Model. In this model, the substrate still needs to fit into the enzyme like a key, but instead of simply fitting into the "keyhole," some type of modification is induced in the substrate, enzyme, or both. The modification begins the process of the reaction. Figure 6-7 illustrates how the Induced-Fit Model applies to the formation of the same enzyme-substrate in Figure 6-6.

All About Kinetics

As you know, all reactions involve energy. The reactants begin with a certain level of energy, an additional quantity of energy is absorbed to reach the transition state (ΔG^*, where the asterisk indicates the transition state), and then energy is released to reach the products. The difference in the energy between the reactants and products is ΔG.

If the energy level of the products is greater than that of the reactants (energy is absorbed), the reaction is *endergonic,* and nonspontaneous. If the energy level of the products is less than the reactants (energy is released), the process is *exergonic,* and spontaneous.

But just because a reaction is spontaneous does not mean it will occur at an appreciable rate. The rate depends on the value of ΔG^*. The greater the value of ΔG^*, the slower the reaction is. An enzyme, like any catalyst, lowers the value of ΔG^* and consequently increases the rate of the reaction. The difference between the reactants and products remains unchanged, as does the equilibrium distribution of the reactants and products. The enzyme facilitates the formation of the transition state (Figure 6-8).

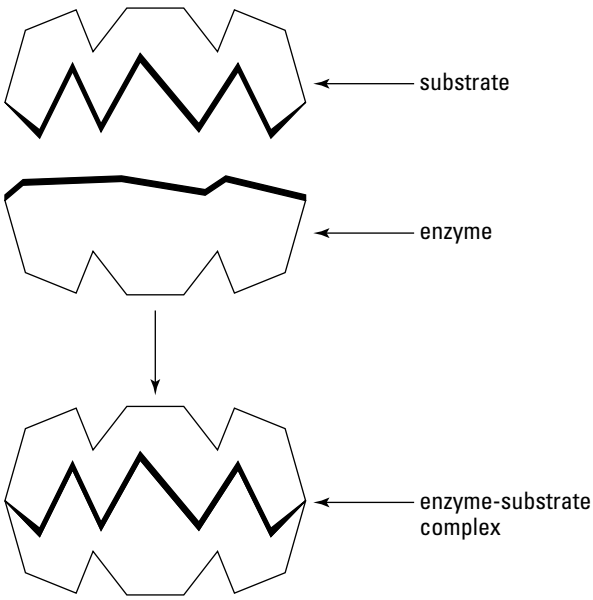


Figure 6-7: The Induced-Fit Model of enzyme catalysis.

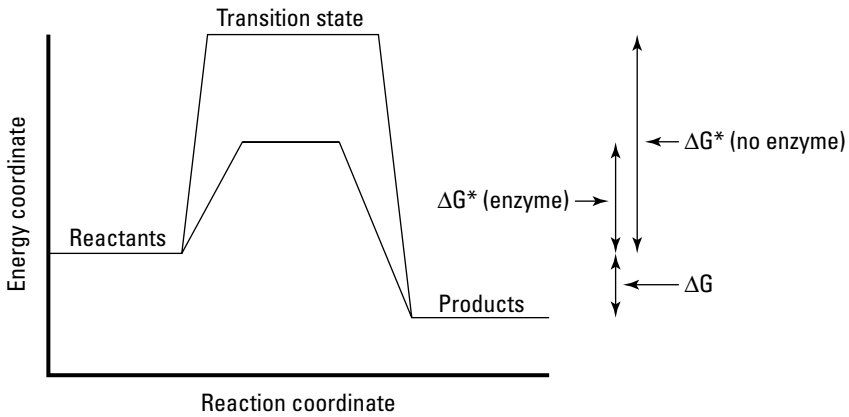


Figure 6-8: Effect of an enzyme on a reaction.

A species has two possible fates in the transition state: It may lose energy and return to the reactant form, or it may lose energy and move to the product form. These two fates lead to two equilibria. One of the equilibria involves the reactant (substrate) and the transition state, and the other involves the product(s) and the transition state. Rapid removal of the product(s) does not allow establishment of the reverse process that leads to the equilibrium. Removal of the product simplifies the analysis of the kinetic data.

Enzymes, like all catalysts, catalyze both the forward and the reverse reaction. The lowering of ΔG^* accelerates both reactions. The ultimate equilibrium concentrations of substrate and products will be the same whether an enzyme is present or not — the enzyme merely changes the amount of time necessary to reach this state.

Enzyme assays: Fixed time and kinetic

An *enzyme assay* is an experiment to determine the catalytic activity of an enzyme. It is possible to measure either the rate of disappearance of the substrate or the rate of appearance of a product. The experimental mode of detection depends on the particular chemical and physical properties of the substrate or the product, and the rate is the change in concentration per change in time. In *fixed time assay,* you simply measure the amount of reaction in a fixed amount of time. In *kinetic assay,* you monitor the progress of a reaction continuously. Once you determine the rate of change in concentration of any reactant or product, it is possible to determine the rate of change of for any other reactant or product of the reaction

It is important to control the conditions precisely. Minor changes in variables such as the temperature or the pH can drastically alter the catalytic activity of an enzyme. For example, the study of enzymes important to humans should be carried out at 37°C, because this is normal body temperature.

Rate determination: How fast is fast?

It is important to control kinetic experiments closely. Once you determine the basic conditions, you can run a series of experiments using a fixed enzyme concentration and varying concentrations of substrate. Up to a point, an increase in substrate concentration results in an increase in rate. The rate increases until the enzyme is saturated. This *saturation point* is where all the enzyme molecules are part of an enzyme-substrate complex. When this occurs, an increase in the substrate concentration yields no increase in the rate, because there are no enzymes available to interact with the

Enzymes in medical diagnosis and treatment

Enzyme levels may indicate medical problems, and that makes enzyme assays useful for both the diagnosis and treatment of medical problems. For examine, creatine kinese (CK) is an enzyme that aids in the synthesis and degradation of creatine phosphate.

CK exists as three different isoenzymes. Each is composed of two polypeptide chains. In the case of muscle CK, the chains are identical, and it's labeled CK-MM. CK found in the brain also has identical polypeptide chains, but they are different from the ones associated with muscle CK and are labeled CK-BB. Finally, the CK found in the heart is a hybrid of the two with one M chain and one B chain: CK-MB.

Normal blood serum contains a little CK-MM and almost no CK-BB and CK-MB. When tissue undergoes injury, though, some of the intracellular enzymes leak into the blood where they can be measured. Elevated levels of total CK (all three isoenzymes) may be indicative of sketal-muscle trauma or myocardial infarction (MI, or heart attack). Analysis of the individual isoenzymes may give additional clues.

For example, an individual falls off a ladder and suffers several broken bones. He is taken to the hospital, where his blood serum CK is measured. It is elevated as expected, but the physician also orders a CK-MB level determination. It turns out to also be highly elevated, indicating that the reason the man fell off the latter to begin with was that he was suffering a heart attack (CK-MB). This knowledge allows the doctor to start a regime of treatment that helps to minimize permanent heart damage.

additional substrate molecules. For most reactions, the rate of the reaction approaches the saturation level along a hyperbolic curve. Theoretically, the reaction rate will only reach saturation at infinite substrate concentration.

A plot of the reaction rate, V, versus the substrate concentration, [substrate], supplies several bits of useful data (see Figure 6-9). The experiment is at constant enzyme concentration. One piece of useful data is the maximum reaction rate, V_{max}. The rate approaches V_{max} asymptotically. At low substrate concentrations, the reaction approaches first-order kinetics, where the rate of reaction depends only on the concentration of one reactant. At high concentrations, the reaction approaches zero-order kinetics, where the rate of reaction is independent of reactant concentration. (Later in this chapter you will see that this graph varies with less simple enzyme-substrate interactions.) In the region between the zero-order region and the first-order region, the kinetics are mixed and difficult to interpret. Important values in the low-concentration region (first-order region) are $\frac{1}{2} V_{max}$ and K_M. The value $\frac{1}{2} V_{max}$ is one-half the V_{max} value. K_M is the Michaelis constant, which corresponds to the substrate concentration producing a rate of $\frac{1}{2} V_{max}$. The Michaelis constant, measured in terms of molarity, is a rough measure of the enzyme-substrate affinity. K_M values vary widely.

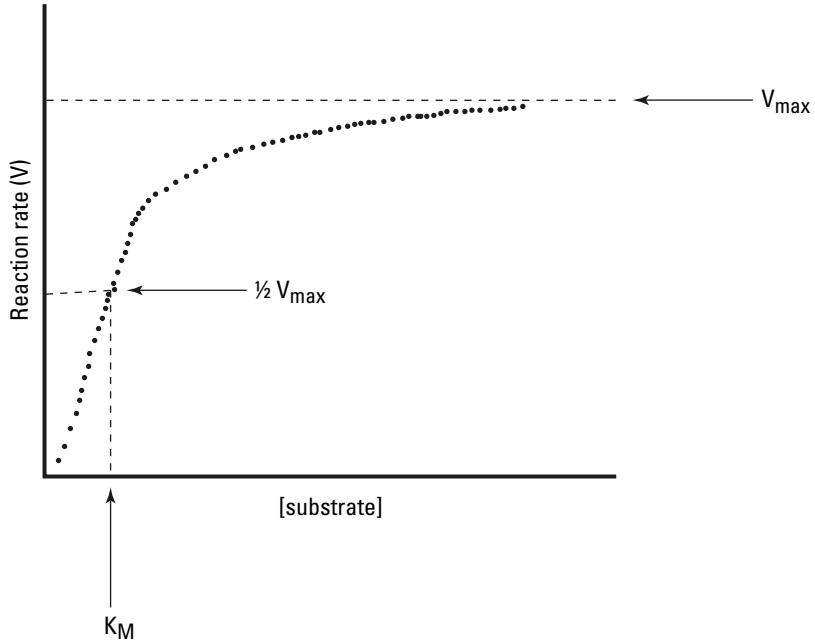


Figure 6-9: Plot of reaction rate, V, versus substrate concentration, [substrate].

At low substrate concentrations, there is an approximately linear relationship between [substrate] and V. At high substrate concentrations, though, V is nearly independent of [substrate]. The low substrate region is useful in the application of the Michaelis-Menten equation (see the next section).

In an uncatalyzed reaction, increasing the substrate concentration does not lead to a limiting V_{max}. The rate continues to increase with increasing substrate concentration. This indirect evidence leads to the conclusion that there is an *enzyme-substrate complex,* a tightly-bound grouping of the enzyme and the substrate. The limit occurs when all the enzyme molecules are part of a complex so that there are no free enzyme molecules available to accommodate the additional substrate molecules. Various x-ray and spectroscopic techniques provide direct evidence to confirm the formation of an enzyme-substrate complex.

Measuring Enzyme Behavior: The Michaelis-Menten Equation

One of the breakthroughs in the study of enzyme kinetics was the development of the Michaelis-Menten equation. It is possible to interpret the behavior of many enzymes by applying the equation to kinetic data. (There are exceptions, and they do not give a graph similar to the one appearing back in

Figure 6-9.) In general, the results of the kinetics experiments are for allosteric enzymes. The *Michaelis-Menten equation* is as follows:

$$V = \frac{V_{max}[S]}{[S] + K_M}$$

In this equation, V is the rate of the reaction, [S] is the substrate concentration, V_{max} is the maximum reaction rate, and K_M is the Michaelis constant. As seen in Figure 6-9, the rate of catalysis, V, increases linearly at low substrate concentration, but begins to level off at higher concentrations. Interpretation begins with examining the following general reaction pathway:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P$$

In this pathway, E refers to the enzyme, S is the substrate, ES is the enzyme-substrate complex, and P is the product. The various instances of k refer to the rate constants of the various steps — a negative rate constant is for the reverse process. In the first step, the separate enzyme and substrate combine to form the enzyme-substrate complex (transition state). The rate of formation of ES is k_1. After ES forms, it may break down to E and S (k_{-1}) or it may proceed to product (k_2). (Note: Some texts refer to k_2 as k_{cat}.)

Because the enzyme will catalyze the reverse process, E and P may combine to reform the complex (k_{-2}). Ignoring the reverse reaction (k_{-2}) simplifies the interpretation of the data. This is not an unreasonable assumption if data collection is near the beginning of the reaction, where the concentration of P is low. The assumption that k_{-2} is negligible leads to a simplification of the preceding equation to:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

Through this simplification, the chemists Leonor Michaelis and Maud Menten were able to propose a model that explains the kinetics of many different enzymes. Through their work, an expression relating the catalytic rate to the concentrations of the enzyme and substrate and to the individual rates was developed. The starting point for this expression is the relationship between the rate of the reaction and the concentration of the enzyme-substrate complex:

$$V = k_2[ES]$$

Similarly, the rate of formation of ES is $k_1[E]\,[S]$, and the rate for the breakdown of ES is $(k_{-1} + k_2)\,[ES]$. Throughout most of the reaction, the concentration of ES remains nearly constant. This is the steady-state assumption, which assumes that during a reaction the concentrations of any intermediates remain nearly constant. This assumption means that the rate of formation of ES must be equal to the rate of breakdown of ES, or:

$$k_1[E][S] = (k_{-1} + k_2)[ES]$$

This equation rearranges to:

$$\frac{[E][S]}{[ES]} = \frac{(k_{-1} + k_2)}{k_1} = K_M$$

The combination of the three rate constants yields a new constant: the Michaelis constant, K_M, which is independent of the enzyme and substrate concentrations and is an important characteristic of enzyme-substrate interactions. Using the Michaelis constant, the concentration of ES is:

$$[ES] = \frac{[E][S]}{K_M}$$

When the enzyme concentration is much lower than the substrate concentration, the value of [S] is very close to the total substrate concentration. The enzyme concentration, [E], is equal to the total enzyme concentration, $[E]_T$, minus the concentration of the enzyme-substrate complex, or $[E] = [E]_T - [ES]$. If we enter this relationship into the preceding equation, we get:

$$[ES] = \frac{([E]_T - [ES])[S]}{K_M}$$

Rearranging this equation gives:

$$[ES] = \frac{[E]_T / K_M}{1 + [S]/K_M} = \frac{[E]_T[S]}{[S] + K_M}$$

Substituting this relationship into $V = k_2[ES]$ or ($V = k_{cat}[ES]$) gives:

$$V = k_2[E]_T \frac{[S]}{[S] + K_M}$$

The maximum rate, V_{max}, occurs when all the enzyme molecules are associated with substrate. That is, $[ES] = [E]_T$. This changes $V = k_2[ES]$ to $V_{max} = k_2[ES]_T$. This relationship changes the preceding equation to the Michaelis-Menten equation:

$$V = \frac{V_{max}[S]}{[S] + K_M}$$

This equation accounts for the information depicted in Figure 6-9. At very low concentrations, $[S] << K_M$, we see $V = (V_{max} / K_M) / [S]$, and when [S] is greater than K_M (high [S]), $V = V_{max}$. When $[S] = K_M$ it leads to $V = V_{max} / 2$.

Ideal applications

The Michaelis-Menten equation explains the behavior of many enzymes. It is relatively easy to determine both the K_M and V_{max} values, and this is normally done graphically using computer programs that generate the best-fit curve.

The K_M values vary widely. The value depends on the identity of the substrate and on a variety of environmental factors such as temperature, ionic strength, and pH. Because K_M indicates the substrate concentration required

to fill half of the active sites on the enzyme, it gives an indication of the minimum substrate concentration for significant catalytic activity to occur. It is possible to determine the fraction of sites filled, f_{ES}, from the value of K_M:

$$f_{ES} = \frac{V}{V_{max}} = \frac{[S]}{[S] + K_M}$$

K_M also gives information about the rate constants for the reaction.

$$\frac{(k_{-1} + k_2)}{k_1} = K_M$$

In the special case where k_{-1} is significantly greater than k_2, $K_M = k_{-1} / k_1$, which relates to the equilibrium constant for the dissociation of the enzyme-substrate complex:

$$K_{ES} = \frac{[E][S]}{[ES]} = \frac{k_{-1}}{k_1}$$

Under these special conditions, K_M is a measure of the binding in the enzyme-substrate complex. A high K_M value indicates that the binding is weak, whereas a low value indicates that the binding is strong.

Don't forget: These conclusions only apply under the special conditions of $k_{-1} >> k_2$.

The value of V_{max} supplies the turnover number of the enzyme. The *turnover number* gives the number of substrate molecules transforming to products per unit of time for a fully saturated enzyme. You can determine k_2 from this value. (The constant k_2 is also known as the catalytic constant, k_{cat}.) If the concentration of active sites, $[E]_T$, is known, this relationship applies:

$$V_{max} = k_2[E]_T$$

And:

$$k_2 = V_{max} / [E]_T$$

Realistic applications

The ideas in the preceding section provide useful information about the behavior of many enzymes. In cells, however, the enzymes are seldom saturated with substrate. Under typical conditions [S] / K_M is usually between 1.0 and 0.01. If K_M is much greater than [S], the catalytic rate k_{cat} (or k_2) is significantly less than the ideal value because only a small portion of the active sites contain substrate. The ratio k_{cat} / K_M allows you to compare the substrate preferences of an enzyme.

The maximum rate of catalytic activity is limited by the rate of diffusion to bring the enzyme and substrate together. Some enzymes can exceed this limit by forming *assemblages.* In these groups, the product of one enzyme is the substrate for a closely associated enzyme. This allows a substrate to enter the group and pass from enzyme to enzyme as if it were in an assembly line.

Another complication is that many enzymes require more than one substrate. It is possible to utilize these multiple substrates through sequential displacement or through double displacement. In *sequential displacement,* all substrates must simultaneously bind to the enzyme before the release of the product. In this type of displacement, the order in which the substrates bind is unimportant. In *double displacement,* or *ping-pong,* situations, one or more products leave before all the substrates bind. Double displacement mechanisms temporarily modify the enzyme.

Here we go again: Lineweaver-Burk plots

Once upon a time, before the invention of computers, the determination of K_M and V_{max} was a tedious process. Today curve-fitting programs allow rapid analysis of the data to determine these values. However, a relatively simple method allows a relatively accurate determination of these two constants. This method is to construct a *Lineweaver-Burk plot,* also known as a *double-reciprocal plot.* The basis of a Lineweaver-Burk plot comes from the manipulation of the Michaelis-Menten equation to the form:

$$\frac{1}{V} = \frac{K_M}{V_{max}} \times \frac{1}{[S]} + \frac{1}{V_{max}}$$

This equation has the form y = mx + b, and describes a straight line. A plot of the reciprocal of the rate, 1 / V, versus the reciprocal of the substrate concentration, 1 / [S], gives a line with a y-intercept equal to 1 / V_{max} and an x-intercept of –1 / K_M. An example of this type of plot appears in Figure 6-10.

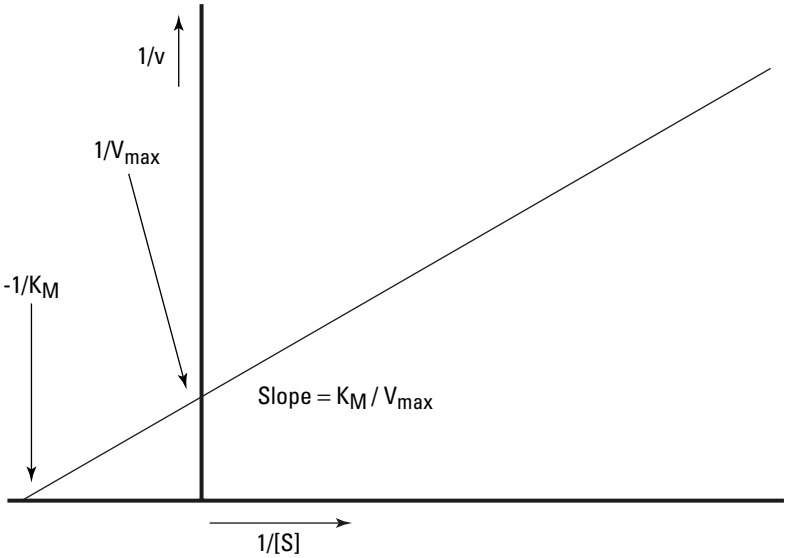


Figure 6-10: A Lineweaver-Burk plot.

The Lineweaver-Burk plot is the most widely used graphical technique for the determination of K_M and V_{max}. However, there are other methods. The Woolf plot, Figure 6-11, uses the equation:

$$\frac{[S]}{V} = \frac{1}{V_{max}} \times [S] + \frac{K_M}{V_{max}}$$

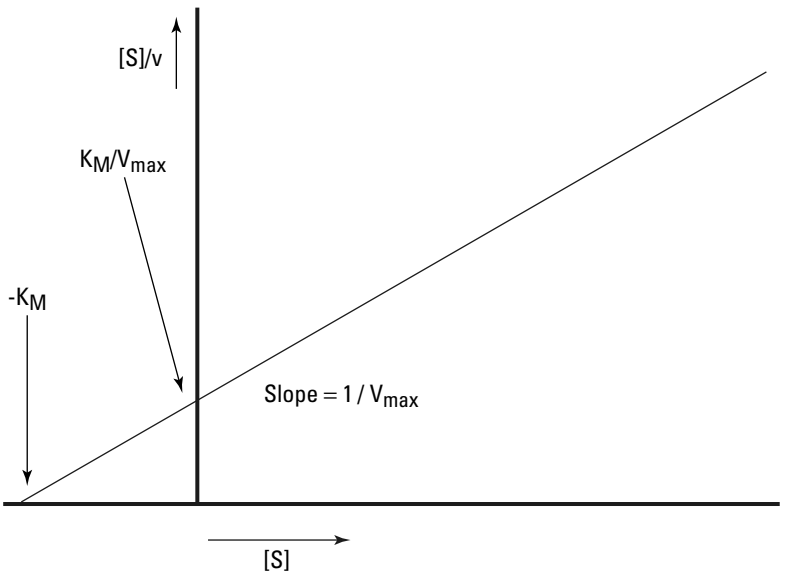


Figure 6-11: A Woolf plot.

Plotting [S] / V versus [S] gives a straight line. An Eadie-Hofstee plot, shown in Figure 6-12, uses the equation:

$$V = -K_M \times \frac{V}{[S]} + V_{max}$$

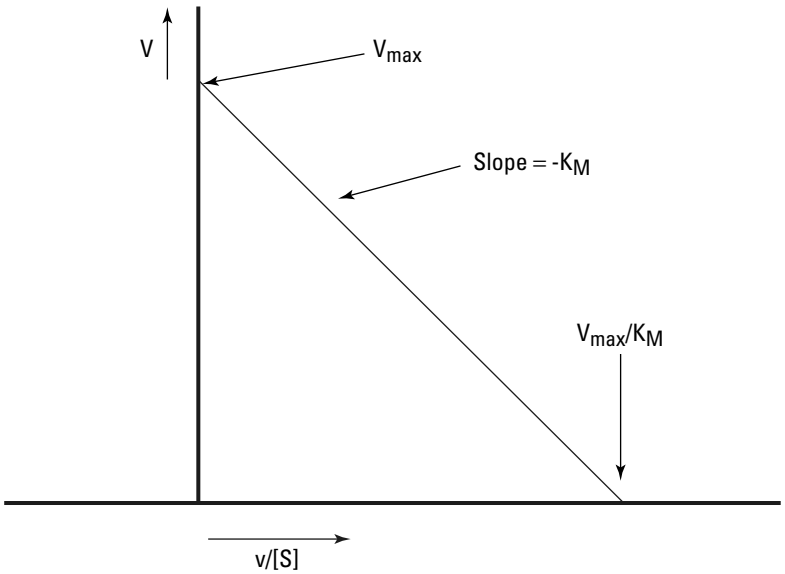


Figure 6-12: An Eadie-Hofstee plot.

Plotting V versus V / [S] gives a straight-line.

Enzyme Inhibition: Slowing It Down

Inibitors are substances that decrease the activity of an enzyme, and they come in two general classes: *competitive* inhibitors, which compete with the substrate, and *noncompetitive* inhibitors, which do not compete. (Mixed inhibition has characteristics of both competitive and noncompetitive inhibition.) In general, these processes are reversible, but there are also irreversible inhibiters that permanently alter the enzyme or bind very strongly to the enzyme. All inhibition may serve as a method of regulating enzymatic activity. There are also many medical applications of this form of inhibition. Examples include anti-epileptic and chemotherapy drugs, along with the ever-popular Viagra. The action of many poisons is also through inhibition.

Competititive inhibition

A competitive inhibitor enters the active site of an enzyme and, thus, prevents the substrate from entering. This prevention results in a decrease in the number of enzyme-substrate complexes that form, and, hence, a decrease in the rate of catalysis. In most cases, a portion of the inhibitor mimics a portion of the substrate. An increase in the substrate concentration overcomes this inhibition because of the increased probability of a substrate molecule entering the active site than an inhibitor molecule.

Noncompetitive inhibition

Noncompetitive inhibitors do not enter the active site but instead bind to some other region of the enzyme. These species usually do not mimic the substrate. This type of inhibitor reduces the turnover number of the enzyme. Unlike competitive inhibition, an increase in the substrate does not overcome noncompetitive inhibition. This type of inhibition takes many different forms, so there is no simple model.

Graphing inhibition

Lineweaver-Burk plots are useful in the study of enzyme inhibition. Figures 6-13 and 6-14 illustrate how the graph changes in the presence of a noncompetitive and a competitive inhibitor. The plot of enzyme inhibition allows us to quickly determine the type of inhibition. In noncompetitive enzyme inhibition, the value of K_M remains unchanged. In competitive inhibition, however, it is $1/V_{max}$ that remains unchanged.

Enzyme Regulation

In general, an increase in the concentration of a substrate, if unregulated, will induce an increase in the rate of reaction. An increase in the concentration of a product will, in general, have the reverse effect. Product regulation is a type of feedback control. In many cases, it is necessary to regulate the activity of enzymes more precisely. There are four general types of enzyme regulation:

- Allosteric control
- Multiple enzyme forms
- Covalent modification
- Proteolytic activation

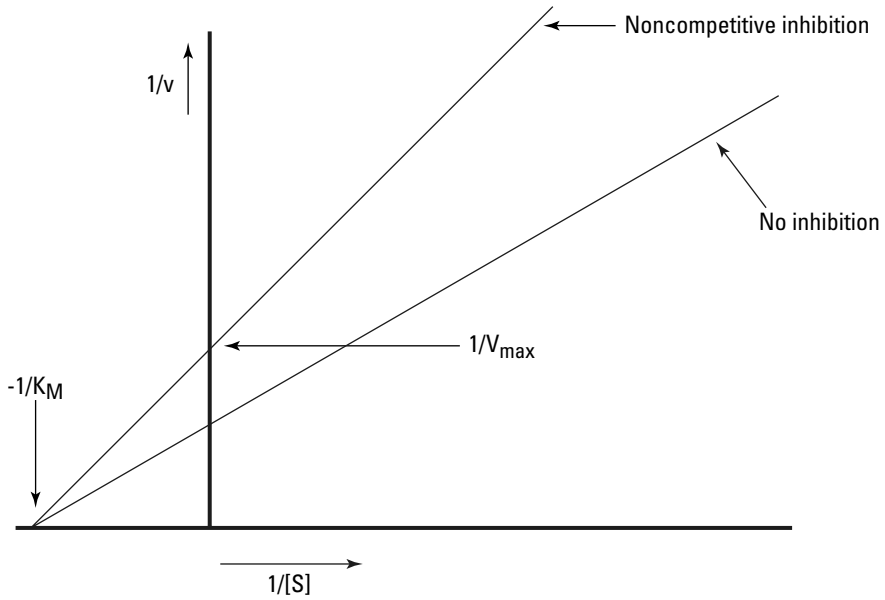


Figure 6-13: A Lineweaver-Burk plot indicating non-competitive inhibition.

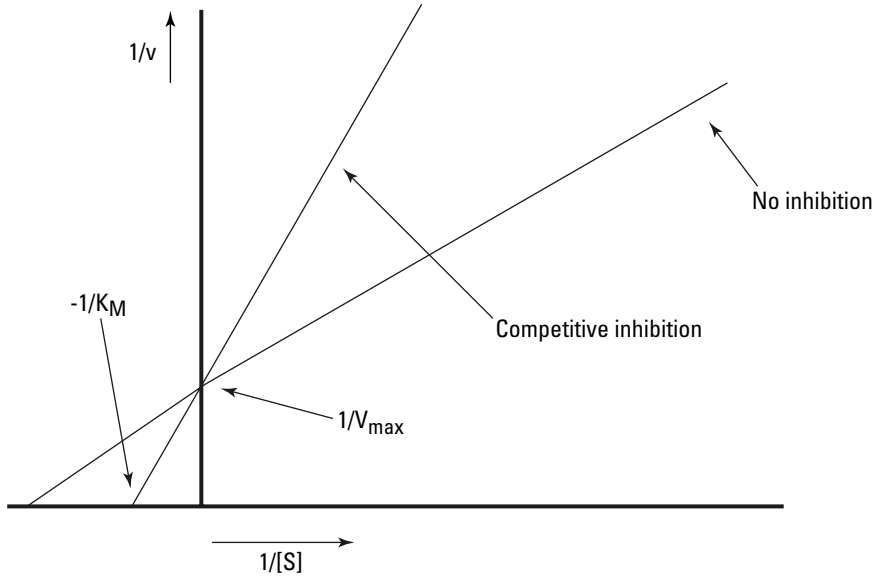


Figure 6-14: A Lineweaver-Burk plot indicating competitive inhibition.

Allosteric control

An allosterically regulated enzyme has a regulatory site. When a small molecule, called a *regulator,* binds to the regulatory site, it induces a conformational change in the enzyme, making it into its active form.

Multiple enzyme forms

Some enzymes have multiple forms known as *isozymes* or *isoenzymes.* There are slight differences in the structures of the forms. These differences lead to differences in the K_M and V_{max} values, and, therefore, in the general activity.

Covalent modification

In this form of regulation, the attachment of a group, often a phosphoryl group, alters the activity of the enzyme. This process is a reversible form of control. Protein kinases catalyze this type of activation, whereas other enzymes catalyze deactivation.

Proteolytic activation

In this form of regulation, an inactive form of an enzyme — a proenzyme or a zymogen — often undergoes irreversible conversion to the active form, often through the hydrolysis of one or more peptide bonds.

Where the money is: Enzymes and industry

The industrial implementation of enzymes originated from studies in the food, wine, and beer industries. Scientists, such as Louis Pasteur, laid much of the groundwork for these applications.

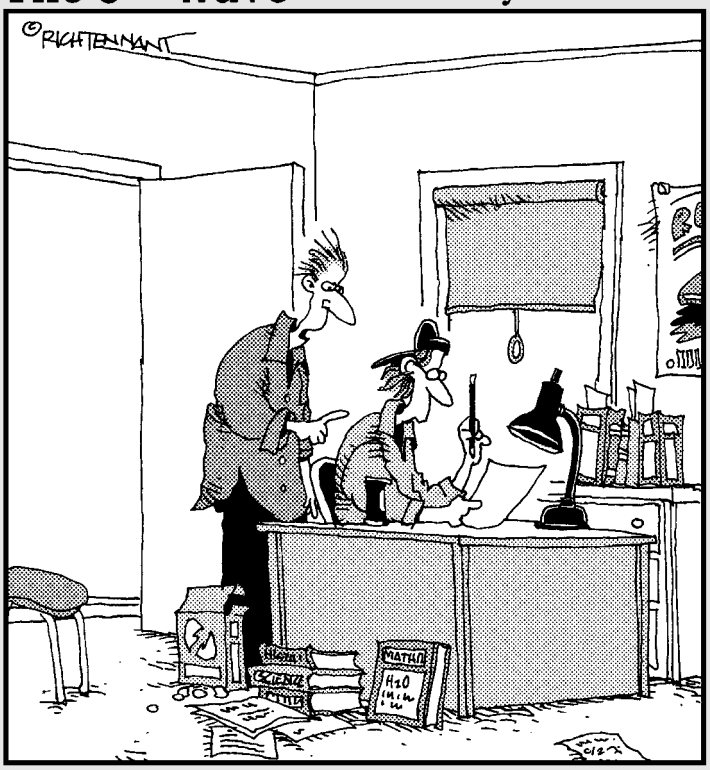
Many of the applications of enzymes to industry involve immobilized enzymes. An *immobilized enzyme* is covalently bonded to an insoluble matrix such as cellulose or glass beads. The immobilization of an enzyme stabilizes it and allows prolonged use. Some useful commercial enzymes are as follows:

Carbohydrases
Amylase: Digestive aid for precooked food
Amyloglucosidase: Converts starch to dextrose
Cellulase and hemicellulase: Conversion of sawdust to sugar and production of liquid coffee concentrates
Glucose isomerase: Production of fructose from cornstarch
Glucose oxidase: Removes glucose from egg solids
Invertase: Stabilizes sugars in soft-centered candy
Lactase: Prevents the crystallization of lactose in ice cream
Pectinase: Clarifies wine and fruit juice
Catalase
Removes H_2O_2 used in the "cold pasteurization" of milk
Proteases
Rennin: Used in cheesemaking
Ficin, Streptodornase, and Trypsin: Debridement of wounds
Pepsin: Digestive aid for precooked food
Papain: Meat tenderizer and beer stabilizer
Bromelain: Meat tenderizer
Alcalase: Additive to detergent for removal of protein stains
Lipoxygenase: Whitening of bread
Lipase: Produces flavor in cheese

Part III

Carbohydrates, Lipids, Nucleic Acids, and More

The 5th Wave By Rich Tennant



"Your formula for a carbohydrate is close, but not entirely accurate. I'm pretty sure carbohydrates consist of carbon, hydrogen, oxygen, sour cream, and bacon bits."

In this part . . .

We go over many biochemical species. Beginning with carbohydrates, we move on to perhaps less tasty-sounding fare: lipids and steroids. Next up: nucleic acids and that amazing encyclopedia about you that sits on the shelf inside every one of your cells: the genetic code of life, guest starring DNA and RNA. After that we end up talking about vitamins and hormones.

Chapter 7

What We Crave: Carbohydrates

In This Chapter

- Finding out about carbohydrates
- Checking out monosaccharides
- Reviewing oligosaccharides

Admit it: You love your carbohydrates. From simple sugars to complex carbohydrates, a day without carbs is a boring day. And carbs are plentiful: In terms of mass, carbohydrates are the most abundant biochemical.

Carbohydrates are a product of photosynthesis, where inorganic carbon dioxide becomes organic carbon with the utilization of solar energy, accompanied by the release of oxygen gas. The conversion of solar energy to chemical energy produces carbohydrates, which are the primary energy source for metabolic processes. Carbs are not only an important energy source but also are the raw materials for the synthesis of other biochemicals. They have structural uses and are a component of nucleic acids.

The term carbohydrate originally referred to "hydrates of carbon" because the general formula of these compounds was $C_nH_{2n}O_n$ or $C_n(H_2O)_n$. However, some materials with this general formula are *not* carbohydrates, and some carbohydrates do not have this general formula. It is better (though not much more conversational) to define carbohydrates as *polyhydroxyaldehydes and polyhydroxyketones and their derivatives*.

Natural carbohydrates are subdivided into *monosaccharides,* or simple sugars containing three to nine carbon atoms, *polysaccharides,* or polymers of monosaccharides, and an intermediate category of *oligosaccharides,* with two to ten monosaccharide units joined. The most important oligosaccharides to humans economically and biologically are the *disaccharrides.*

Properties of Carbohydrates

In general, the names of most carbohydrates are recognizable by an *-ose* suffix. An *aldose,* for example, is a monosaccharide where the carbonyl group is an aldehyde, whereas in a *ketose* the carbonyl group is a ketone. Chemists also use roots referring to the number of carbon atoms. *Pentoses,* five-carbon atoms, and hexoses, six-carbon atoms, are very important. *Trioses, tetroses,* and so on are also found in nature. It is possible to combine these generic names to give terms such as *aldohexose* and *ketopentose.*

They contain one or more chiral carbons

Chiral carbons are those that have four different groups, atoms or groups of atoms, attached to them. Most carbohydrates contain one or more chiral carbons. For this reason, they are *optically active,* rotating polarized light in different directions and many times having different activity in biological systems. Fischer projections are useful in indicating the asymmetry around each of the chiral carbon atoms. Figure 7-1 illustrates the construction of a Fischer projection. In the Fischer projection, the vertical lines project back, and the horizontal lines project forward. There are two arrangements of groups around a chiral center: These arrangements are called *enantiomers* and represent nonsuperimposable mirror images, like left and right gloves. The enantiomers comprise a D/L pair, where the D form rotates polarized light to the right, and the L form rotates polarized light to the left.

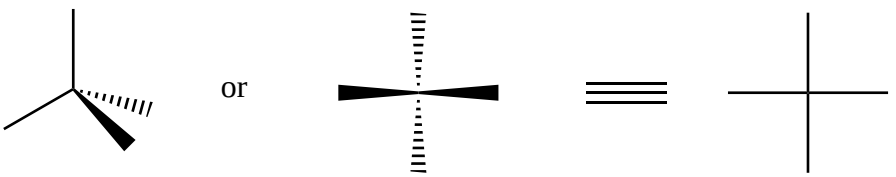


Figure 7-1: The relationship between the three-dimensional structure and the Fischer projection.

Fischer projections are not only useful for representing chiral carbons, but they are useful in identifying which enantiomeric form is present in a sample. To determine whether two projections are enantiomers or just simply two representations of the same molecule, it is necessary to compare the

two drawings. During this comparison, rotate one of the projections by 180° about an axis perpendicular to the plane of the paper (in other words, turn the paper while it's lying on a table). If the diagrams are identical after this rotation, then they are simply two representations of the same molecule. If the diagrams are not identical, they represent a pair of enantiomers.

They have multiple chiral centers

Because many carbohydrates have more than one chiral center (more than one chiral carbon), there can be more than two stereoisomers. The number of stereoisomers is 2^n, where n is the number of chiral carbons. For example, if the compound has two chiral carbons, there are a total of four stereoisomers — two pairs of enantiomers. Although the members of each pair are enantiomers, members of the different pairs are referred to as *diastereomers*.

The structure for D-glucose, a typical monosaccharide, appears in Figure 7-2. In this figure (a Fischer projection), all the carbon atoms except the ones at the top and bottom are chiral — a common way of representing monosaccharides. The carbon atoms appear as a vertical chain with the carbonyl carbon as near the top as possible (it is at the top for an aldose). Numbering the carbon atoms begins at the top, as indicated with the top carbon labeled C_1. The highest-numbered chiral carbon in this case is number five. By convention, if the –OH on this carbon atom appears on the right, it's the D form of the monosaccharide; if it is on the left, it's the L form.

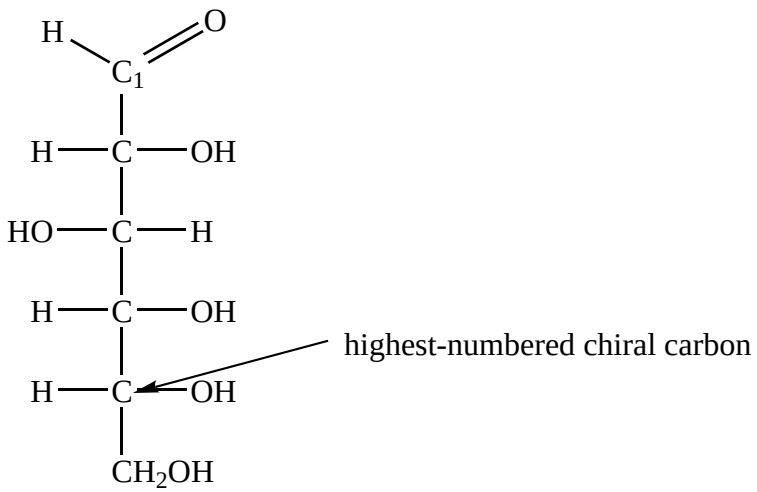


Figure 7-2: Structure of D-glucose.

Any change in the relative positions of the groups attached about any of the chiral carbon atoms in a Fischer projection produces either a different enantiomer or a diastereomer (assuming that the result is not simply a different way of drawing the original structure). In the case of D-glucose, with 4 chiral centers, there are 16 structures. One is D-glucose, and another is its enantiomer: L-glucose. The remaining 14 structures are diastereomers consisting of 7 enantiomeric pairs. Each of the enantiomeric pairs consists of a different monosaccharide. In the case of glucose, you have glucose, allose, altrose, mannose, gulose, idose, galactose, and talose, shown in Figure 7-3. The different D-ketohexoses are in Figure 7-4.

A Sweet Topic: Monosaccharides

The *monosaccharides,* or simple sugars, are an important class of biochemicals. For example, glucose, one of the most common monosaccharides, is the primary form of energy storage in the body. Most monosaccharides taste sweet. The relatively large number of hydroxyl groups and the polar carbonyl group mean that most of these compounds are water-soluble. And, as mentioned earlier, most are optically active.

The most stable monosaccharide structures: Pyranose and furanose forms

The most important monosaccharide is D-glucose (one form of D-glucose appears back in Figure 7-2). This form exists in equilibrium with two slightly different ring forms. The ring form results from an internal cyclization reaction, where a two groups on the same molecule join forming a ring. (The rings appear as planar structures even though the actual structures are not planar.) This cyclization involves a reaction between the carbonyl group and the highest-numbered chiral carbon, producing one of the following structures: a hemiacetal, an acetal, a hemiketal, or a ketal. In the case of D-glucose a pyranose ring forms. Haworth projection formulas are useful when representing the ring forms of a monosaccharide (Figure 7-5).

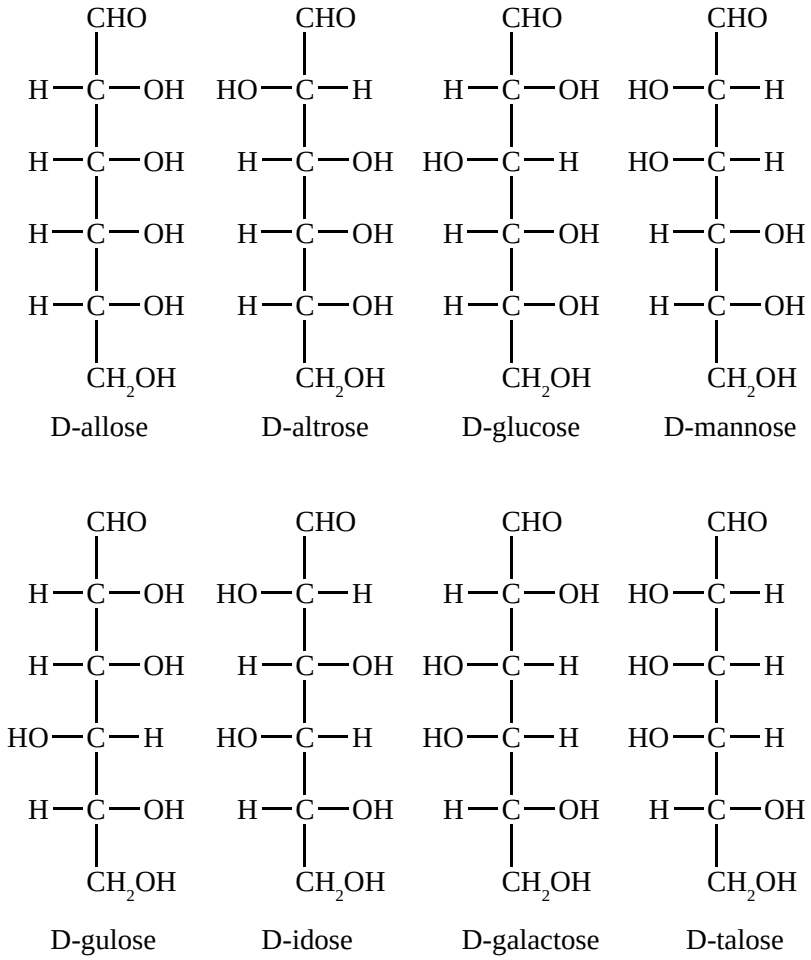


Figure 7-3: Structures of the D-aldohexoses.

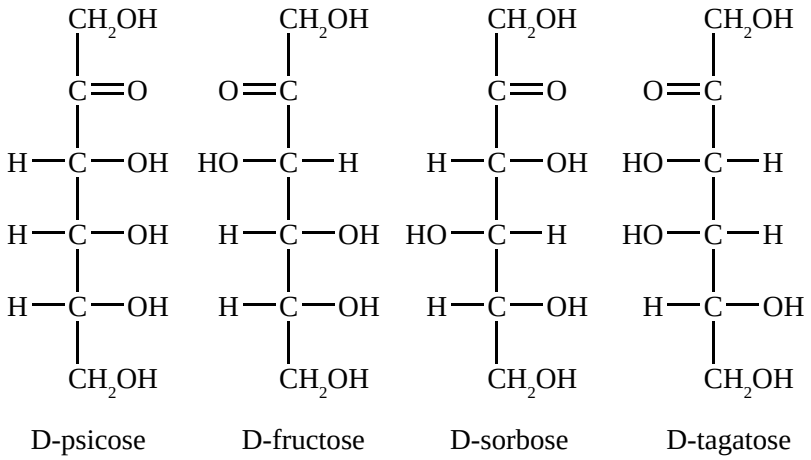


Figure 7-4: Structures of the D-ketohexoses.

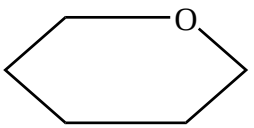


Figure 7-5: A pyranose ring.

There are two possible structures for the pyranose structure of D-glucose (see Figure 7-6), and other monosaccharides. If we examine the Fischer projection for D-glucose, we can see why:

- **Structure 1:** Hydroxyl group on one carbon in the up position.
- **Structure 2:** Hydroxyl group on the corresponding carbon in the down position.

If you "bend" the carbonyl group around and then allow a reaction with the highest numbered chiral carbon, you have two choices: right or left. This gives two forms known as *anomers.* The anomers are labeled α and β. The carbonyl carbon — C_1, in this case – is the anomeric carbon, which should be on the right side of a Haworth projection. The relative positions of –H and –OH about the anomeric carbon determine whether it is the α or β form. The hydroxyl group points down in the α form, and the hydroxyl group points up in the β form. (Reversing the drawing of the rings may give a structure with the opposite orientation of the groups about the anomeric carbon.) In solution, each of the anomers is in equilibrium with the open chain form represented by the Fischer projection. Therefore, there is an interconversion between the α and β forms known as *mutarotation.*

It is also possible to form a five-membered ring, called a *furanose* ring. A simplified furanose structure appears in Figure 7-7. Ribose is an example of a monosaccharide that may form a furanose ring.

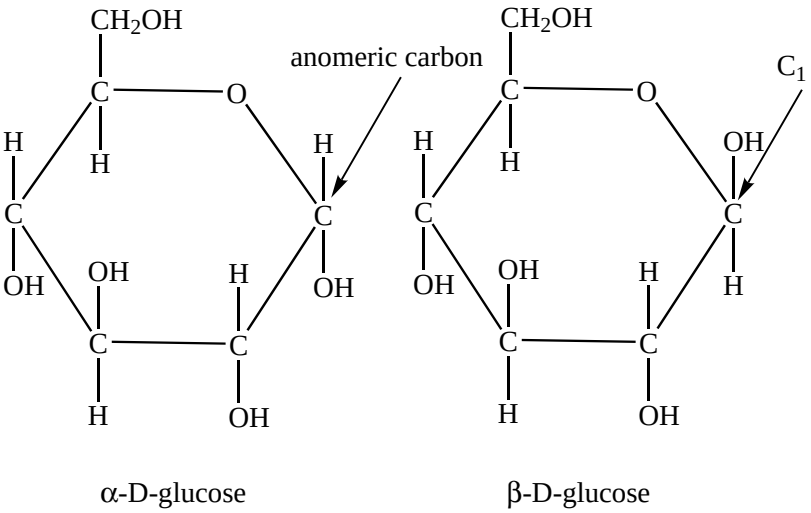


Figure 7-6: The Haworth projections for the pyranose structures of D-glucose.

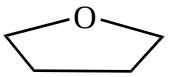


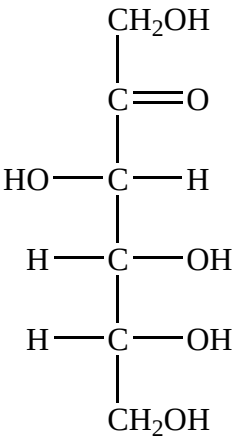
Figure 7-7: A furanose ring.

The pyranose and furanose forms are the thermodynamically more stable forms of the monosaccharides. In general, in the equilibria involving ring and open forms, less than ten percent of the molecules are in the open form. Fructose is a ketose that may form a furanose ring. Structures of D-fructose are shown in Figure 7-8.

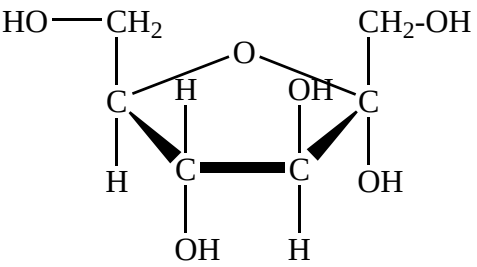
Chemical properties of monosaccharides

Many aldoses, because of the aldehyde group, are reducing sugars — that is, they are reducing agents in certain redox reactions. A number of tests for reducing sugars, include using Fehling's solution or Benedict's solution. These tests are useful to check for glucose in the urine of a diabetic.

The reaction of a monosaccharide with methanol, CH_3OH, in the presence of hydrochloric acid, HCl, replaces the hydrogen atom of the hydroxyl group on C_1 with a methyl group, forming a *glycosidic bond.* (Nitrogen may also be part of a glycosidic bond.) Once the glycoside forms, the ring is "locked," meaning it will not reopen; therefore, mutarotation will no longer take place. A formerly reducing sugar will no longer be a reducing sugar.



D-fructose



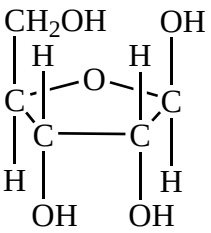
α-D-fructose

Figure 7-8: Two forms of D-fructose.

Derivatives of the monosaccharides

A variety of derivatives of the monosaccharides are formed through the alteration of one or more of the functional groups present. In this section we examine some of these derivatives using D-ribose as the parent monosaccharide. Two forms of the structure of D-ribose appear in Figure 7-9.

The reduction of the carbonyl group to an alcohol yields a reduced sugar (polyhydric alcohol). The reduction of D-ribose forms D-ribitol (Figure 7-10).



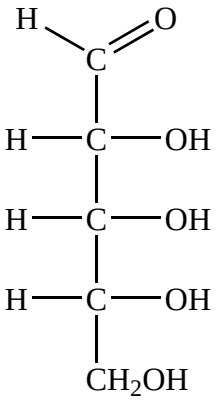


Figure 7-9: Two representations of the structure of D-ribose.

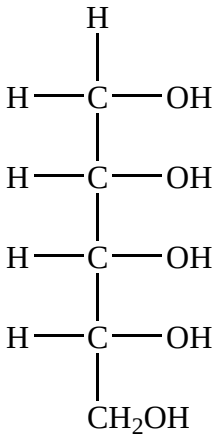


Figure 7-10: D-ribitol.

It is also possible to oxidize a monosaccharide to a carboxylic acid. There are two important oxidations: oxidation of an aldehyde (aldose) to an aldonic acid, and oxidation of the alcohol on the highest-numbered carbon atom to a uronic acid. In the case of D-ribose, it is possible to form D-ribonic acid (Figure 7-11) or D-ribouronic acid (Figure 7-12).

Monosaccharides, like all alcohols, may react with acids to form esters. The combination with phosphoric acid (phosphate sugar) is a biologically important reaction. Any of the alcohol groups may react. Figure 7-13 shows one example: D-ribose-1-phosphate. (The "1" refers to the attachment of the phosphate group to C_1.)

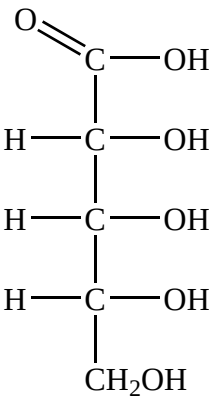


Figure 7-11: D-ribonic acid, an aldonic acid.

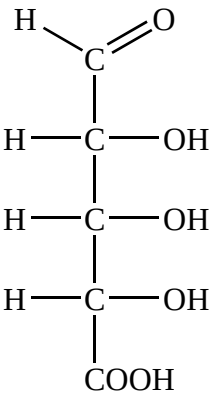


Figure 7-12: D-ribouronic acid, a uronic acid.

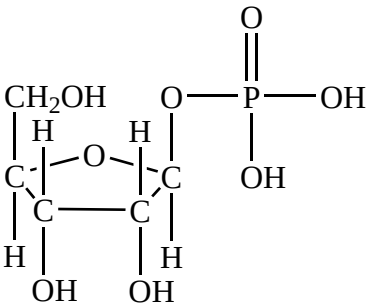


Figure 7-13: D-ribose-1-phosphate.

The most common monosaccharides

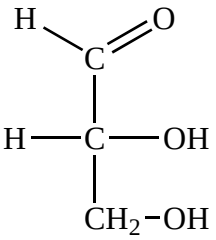
Glucose, or blood sugar, is also known as dextrose. The anomeric carbon is part of a hemiacetal, and the name of the pyranose structure is *glucopyranose.*

Blood is commonly tested for blood glucose levels, which are controlled by the hormone *insulin,* produced within the body in the pancreas. In a healthy human, blood glucose levels rise slightly after eating. The pancreas then releases insulin in order to keep the levels from rising too high. A healthy individual has a fasting blood sugar of 70–99 milligrams of glucose per deciliter of blood and 70–145 mg/dL two hours after eating. The American Diabetes Association associates blood glucose levels of 126 mg/dL (fasting) or 200 mg/dL (two hours after eating) with *diabetes* — the inability of the pancreas to produce enough insulin.

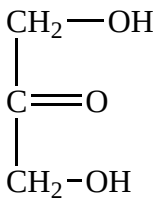
The simplest aldose is glyceraldehyde, and the simplest ketose is dihydroxy acetone. Figure 7-14 shows the structures of these two compounds.

The beginning of life: Ribose and deoxyribose

The monosaccharides D-ribose and D-deoxyribose are important components of the nucleic acids. They are present in these complex molecules in the form of a furanose ring. In addition, they are present as the β anomer. The difference between these two monosaccharides is that there is one less oxygen atom present in deoxyribose, hence the "deoxy." The "missing" oxygen atom is at C_2. The structures of these two sugars appear in Figure 7-15.

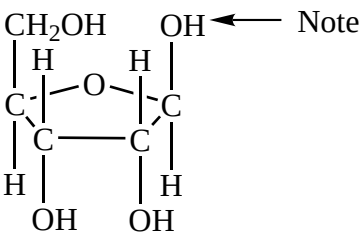


D-glyceraldehyde

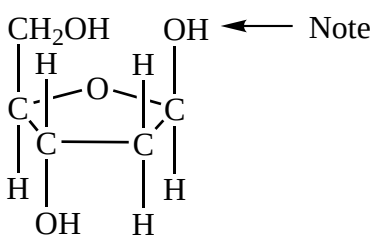


Dihydroxyacetone

Figure 7-14: Glyceraldehyde and dihydroxyacetone.



ribose



deoxyribose

Figure 7-15: The arrows point to the positions of the alcohol groups leading to these becoming the β anomers.

Sugars Joining Hands: Oligosaccharides

The joining of two or more monosaccharides forms an *oligosaccharide,* with two to ten monosaccharide units, or a *polysaccharide,* a polymer having many more monosaccharide units. One or more glycoside linkages hold the monosaccharides together. The simplest, and most common, oligosaccharides are the disaccharides.

Keeping it simple: Disaccharides

A *disaccharide* is an oligosaccharide composed of two monosaccharide units. The best-known disaccharide (and surely the most well liked) is probably sucrose, which you know as table sugar or cane sugar. Each molecule of this sugar is a combination of a glucose molecule and a fructose molecule. There are many other important disaccharides — among them, maltose, malt sugar, and lactose, milk sugar, each of which contains two molecules of glucose. Due to its simplicity,where two identical monosaccharides are joined, we will use maltose to illustrate several points concerning disaccharides, and, by implication, other oligosaccharides and polysaccharides. The structure of maltose appears in Figure 7-16.

The oxygen atom joining the two glucose rings of the maltose molecule in Figure 7-16 is a glycoside linkage — an α(1-4) linkage. The α refers to the anomeric form of the ring on the left. If β-D-glucose were present instead, then lactose would result (see Figure 7-17). The 1-4 indicates that C_1 of the left ring links to C_4 of the right ring. The loss of a water molecule accompanies the formation of the linkage, which locks the left ring so that mutarotation is no longer possible. The locked ring is also no longer a reducing sugar. But mutarotation can still occur on the right ring.

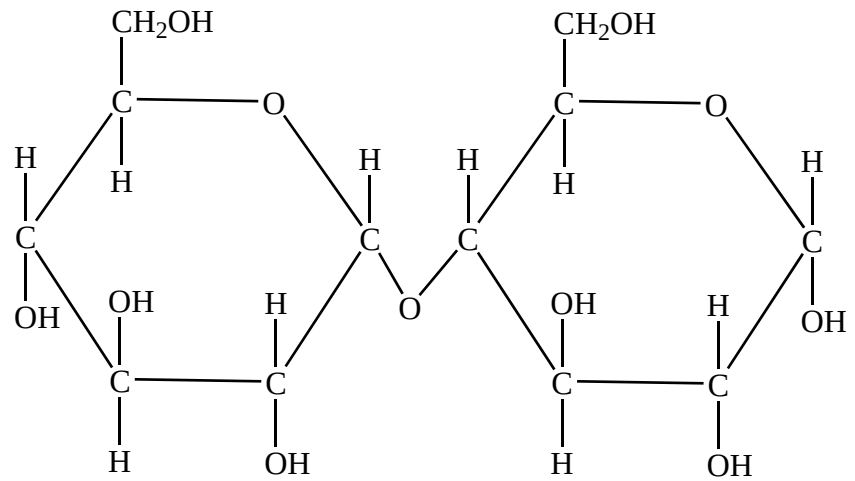


Figure 7-16: The structure of maltose with a α(1-4) linkage present.

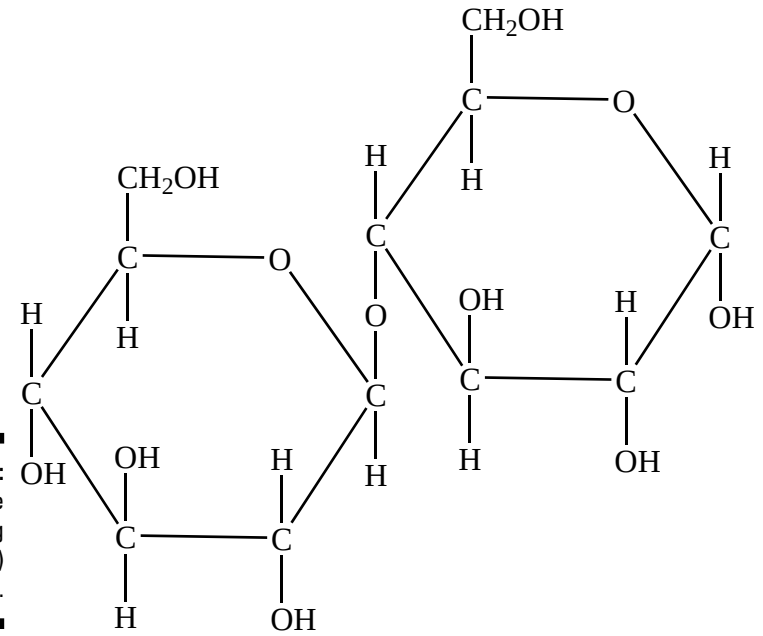


Figure 7-17: Lactose showing its β(1-4) linkage.

Sucrose, table sugar, is a disaccharide like maltose. It forms when D-glucose links to a D-fructose by a α(1-2) linkage. This situation locks both rings so that mutarotation of neither ring can occur. The formation of sucrose appears in Figure 7-18.

If the sweetness of sucrose is 100, then the sweetness level of glucose is 74, and that of fructose, 173. Fructose, found in corn syrup, is the sweetest common sugar — meaning you need less of it to make foods taste sweet. Less sugar translates to fewer calories. There are also naturally occurring, sweet-tasting proteins, some of which are hundreds of times sweeter than sugar.

Quite a few artificial sweeteners are used in commercial products. The best known are *saccharin* (about 500 times as sweet as sucrose, *aspartame* (200 times as sweet as sucrose), and *sucralose* (marketed as Splenda) — which is a whopping 600 times as sweet as sucrose. Sucralose is created by replacing three of the hydroxyl groups of sucrose with chlorines.

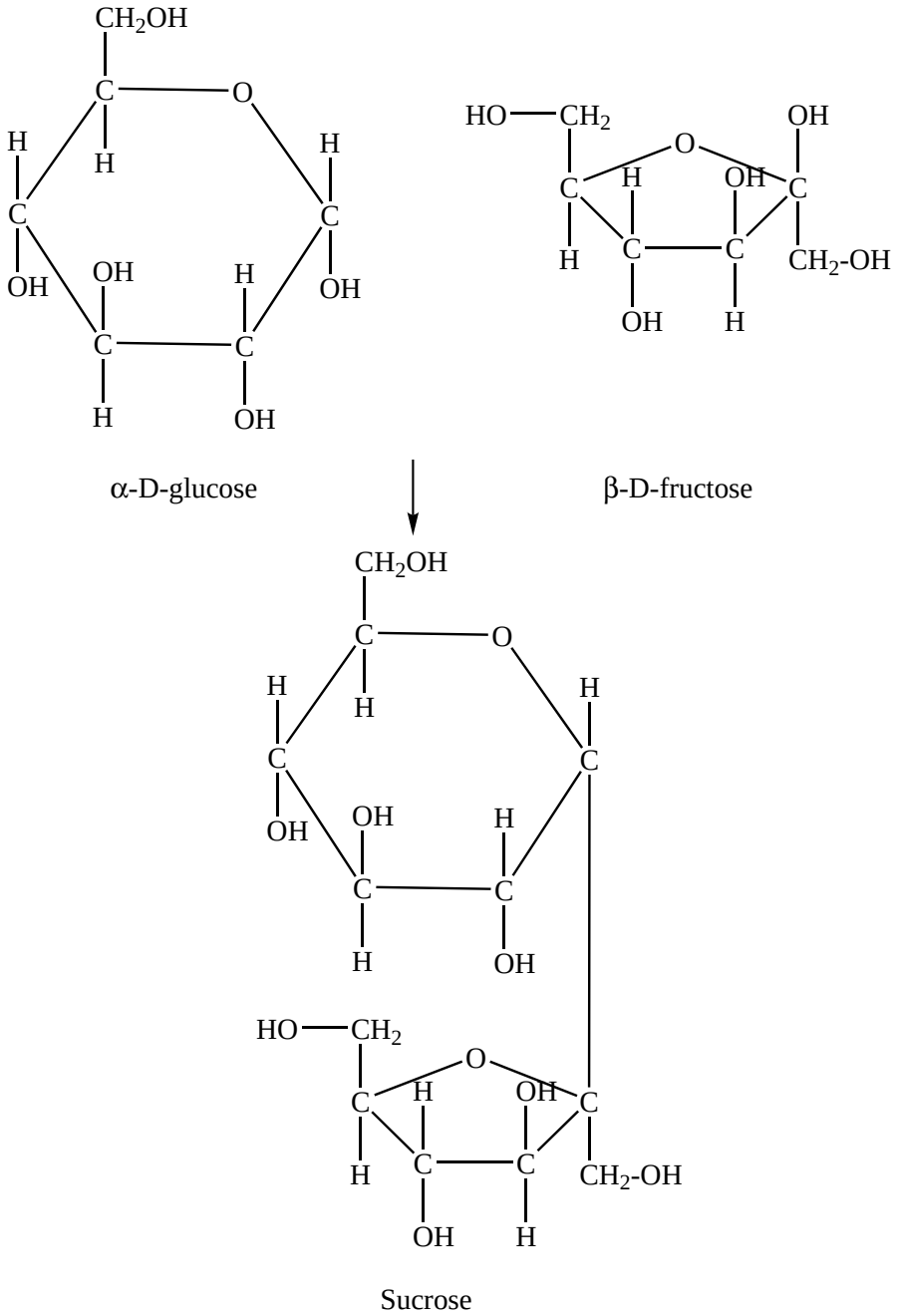


Figure 7-18: Structure of sucrose, formed by joining α-D-glucose and β-D-fructose.

Starch and cellulose: Polysaccharides

The two most important polysaccharides are starch and cellulose. Both of these are polymers of D-glucose. The basic difference between these two polymers is the linkages between the glucose units. Starch is related to maltose and cellulose is related to lactose.

Bread, pasta, and potatoes: Starches

Of all the carbohydrates, we think starches are our favorite. Bring on the potatoes and pastas! The different types of these lovely, delicious polysaccharides are very closely related by the linkages between their monomer units. *Starch* is a polymer of α-D-glucose. There are three common types of starch: amylase, amylopectin, and glycogen. Amylase is the combination of α(1-4) glucose groups. Amylopectin, like amylase, has α((1-4) glucose linkages, but, in addition, it has α((1-6) branches. Glycogen, animal starch, is similar to amylopectin except that it has more branches. All three are useful in storing glucose, and all three give an intense dark blue color in the presence of iodine — a simple and useful test.

Keeping the termites happy: Cellulose

Ever wonder why you can eat a potato but not a tree? *Cellulose* is similar to starch except that the linkages are β(1-4) glucose. The primary use of cellulose in nature is structure. Cleavage of the linkages is only possible with enzymes produced by certain bacteria or fungi. For this reason, only certain creatures, such as termites, and ruminants like cows, who have these bacteria in their GI tracts, can digest and utilize cellulose as an energy source. Cellulose is one of the most abundant biochemicals on earth.

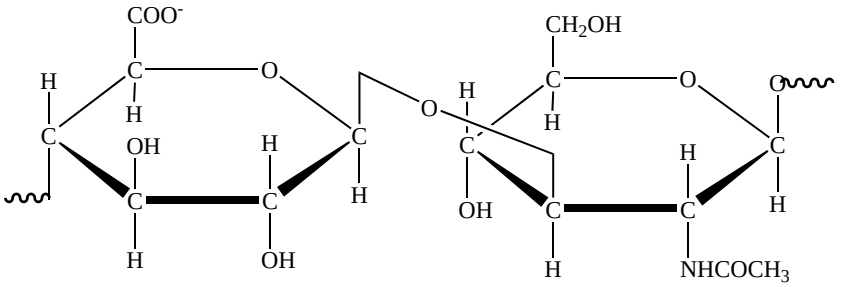
Biological connective tissue: Acidic polysaccharides

One of the major uses of polysaccharies in the body is the area of connective tissues, the compounds that hold our parts together. This group of tissue includes tendons, ligaments, and collagen. (Fuller lips, anyone?) Acidic polysaccharides are important to the structure and function of connective tissue. The repeating units of these polysaccharide derivatives are disaccharides. One of the components of the disaccharide is an *amino sugar* (where an amino group substitutes for an alcohol group). One or both of the components of the disaccharide unit have a negatively charged group (either a sulfate or a carboxylate). Examples are hyaluronic acid and heparin. The hyaluronate and heparin repeating units appear in Figure 7-19.

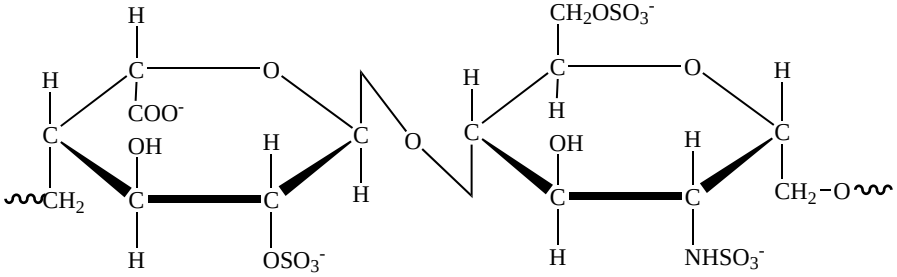
Heparin is used to treat and prevent blood clots from forming, especially in the lungs and legs. It is commonly used after dialysis, after surgery or when the patient has been unable to move for extended periods of time. It acts as an anticoagulant by binding to one of the anti-clotting proteins, increasing its efficiency up to a thousand-fold.

Glycoproteins

Most of the proteins occurring in blood serum are *glycoproteins,* which are proteins with carbohydrates attached. The presence of the carbohydrate tends to increase the hydrophilic nature of the protein. In general, the linkage is by attachment to an asparagine, serine, or threonine residue. Some soluble proteins and some membrane proteins are glycoproteins. We will see glycoproteins again at various times in later chapters.



Hyaluronate



Heparin

Figure 7-19: Disaccharide repeating units in hyaluronate and heparin.

Chapter 8

Lipids and Membranes

In This Chapter

- Living with lipids
- Examining triglycerides
- Finding out about membranes
- Seeing how steroids and other lipids operate

Along with cholesterol, lipids tend to have a bad reputation in today's world, even though they are absolutely necessary to good health. The *lipids* are an exceedingly diverse group of biologically important materials that are distinguished by solubility. A lipid is a member of a group of compounds that are not soluble (or only sparingly soluble) in water but that are soluble in nonpolar solvents or solvents of low polarity. The nonpolar nature of lipids is due to the fact that a large portion of the molecule contains only carbon and hydrogen. If there were significant amounts of oxygen or nitrogen in the structure, the substance would be more polar and hence more soluble in water.

Lovely Lipids: An Overview

Lipids have many important biological roles, including being highly concentrated energy sources, membrane components, and molecular signals. There are lots of kinds of lipids. Figure 8-1 provides a diagram showing the relationship between many of the different categories of lipids. Arachidonic acid, a fatty acid, appears in Figure 8-1 twice — once as the precursor (compound leading) to leukotrienes and prostaglandins and again as a member of the fatty acid group. We double-listed arachidonic acid this way because of its very different roles in these two chemical pathways.

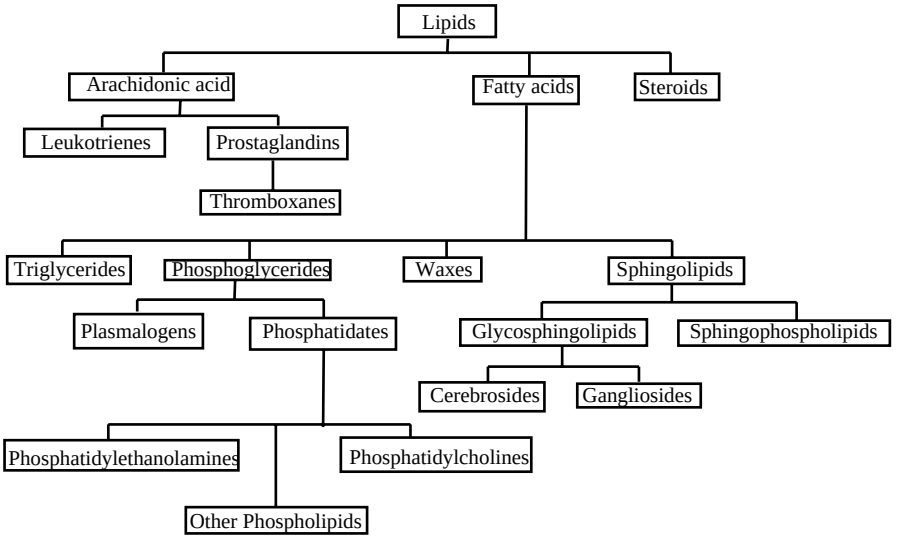


Figure 8-1: The relationships among the many types of lipids.

In the body, lipids provide energy storage and structure (cell membranes) and regulate bodily functions. Many of the lipids work like soap and detergents. Like soaps, lipids have a nonpolar region — usually a fatty acid — and a polar region. Figure 8-2 shows a representation of the structure of a soap.

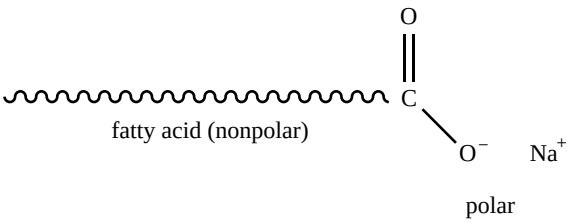


Figure 8-2: Representation of a soap.

In water, soap forms a *micelle,* (see Figure 2-3 in Chapter 2) in which the nonpolar portions of the different molecules coalesce and leave the polar portions on the outside next to the water. If there is any other nonpolar material present, such as grease from dirty dishes, it tends to migrate to the interior of the micelle. With the polar portions of the soap molecules on the outside, the micelle appears as one large polar molecule instead of a number of smaller molecules that have polar and nonpolar regions.

The *dual solubility* nature of soap is why it removes grease or oil from your skin or clothes. The grease or oil is nonpolar and, therefore, is not soluble in water. The soap forms a micelle that surrounds the grease/oil in the nonpolar portion of the micelle. The polar end of the soap micelle is soluble in water,

allowing the grease and oil to be removed during rinsing. Although many different types of lipids exist, our discussion in this chapter focuses on the following four types of lipids:

- **Fatty acids and derivatives (esters):** Fats, oils, and waxes
- **Complex lipids:** Phosphoglycerides and sphingolipids
- **Steroids**
- **Arachidonic acid devivatives:** Prostaglandins, thromboxanes, and eukotrienes

Lipids are important not only as individual molecules but also in terms of their interactions with other lipids and non-lipids in the formation of lipid bilayers or cell membranes. These interactions occur both at the cell boundary and around some interior structures. The fatty acids portions of the lipids are especially important in their physical and chemical properties. The naturally occurring fatty acids have a few key features:

- They are all straight-chained with generally 10–20 (but sometimes more) carbon atoms.
- They have an even number of carbon atoms.
- If carbon-carbon double bonds are present, only the cis-isomer is present.

Table 8-1 lists a few of the common fatty acids.

Table 8-1	Common Fatty Acids
Lauric acid	$CH_3(CH_2)_{10}COOH$
Myristic acid	$CH_3(CH_2)_{12}COOH$
Palmitic acid	$CH_3(CH_2)_{14}COOH$
Palmitoleic acid	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COOH$
Stearic acid	$CH_3(CH_2)_{16}COOH$
Oleic acid	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$
Linoleic acid	$CH_3(CH_2)_3(CH_2CH{=}CH)_2(CH_2)_7COOH$
Linolinic acid	$CH_3(CH_2CH{=}CH)_3(CH_2)_7COOH$
Arachidonic acid	$CH_3(CH_2)_4(CH{=}CHCH_2)_4(CH_2)_2COOH$

A *wax* is a simple ester of a fatty acid and a long-chain alcohol. The fatty acid typically contains at least 10 carbon atoms, whereas the alcohol portion is typically 16–30 carbon atoms. In general, a wax, such as the wax in your ears, serves as a protective coating. Because they tend to be somewhat unreactive, we do not discuss waxes in much detail in this book.

A Fatty Subject: Triglycerides

Fats (and oils) are *triglycerides* or *triacylglycerols.* That is, they are triesters of fatty acids with glycerol. *Glycerol* is a trihydroxy alcohol (see Figure 8-3). In a fat, each of the three alcohol groups becomes part of an ester through the reaction with a fatty acid. The fatty acids may or may not be the same.

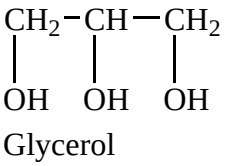


Figure 8-3: Structure of glycerol.

Properties and structures of fats

The basic difference between a fat and an oil is that a fat is a solid at room temperature and an oil is a liquid. That said, two important structural factors distinguish a fat from an oil. One is the size of the fatty acids, and the other is the presence or absence of double bonds. The longer the fatty acid chain, the higher the melting point. The greater the number of carbon-carbon double bonds, the lower the melting point.

A *saturated* fat consists of fatty acids with no carbon-carbon double bonds. An *unsaturated* fat has a double bond while a *polyunsaturated* fat has multiple double bonds.

There are more than 70 known naturally occurring fatty acids. In most natural fats, there is a cis geometry about the double bonds. The presence of double bonds puts "kinks" in the carbon chain, which prevent the fatty acid chains from stacking together as roughly parallel chains. The inability of unsaturated fatty acid chains to stack together inhibits the fat's ability to solidify.

The treatment of an unsaturated fat or oil with hydrogen in the presence of a catalyst such as nickel will lead to hydrogenation of some or all of the carbon-carbon double bonds, forming carbon-to-carbon single bonds. This procedure changes an unsaturated fat into a saturated fat. In most cases, only

partial hydrogenation takes place, and the hydrogenation raises the melting point of the compound. By this procedure, it is possible to convert an oil (liquid) into a fat (solid). Incomplete hydrogenation may change some of the cis arrangements into trans arrangements, producing a *transfat.*

Figure 8-4 shows the structure of a typical fat. Note that that the two upper fatty acid chains (saturated) "stack" next to each other, but the lower chain (unsaturated) does not.

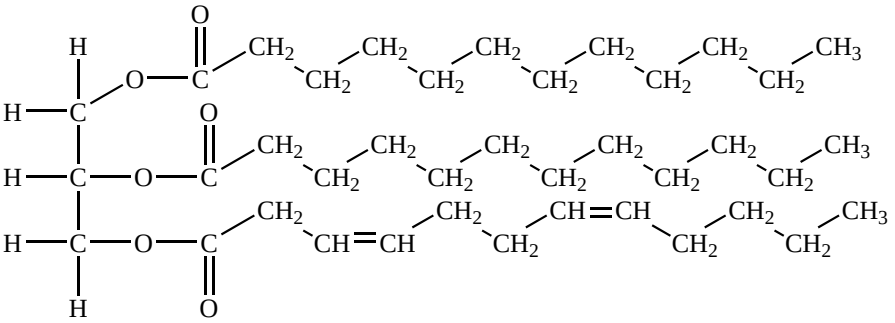


Figure 8-4: Structure of a typical fat: Upper chains are saturated, bottom chain is unsaturated.

Cleaning up: Breaking down a triglyceride

For centuries, the treatment of a fat (commonly animal fat) with a strong base catalyst (generally lye — sodium hydroxide) has been used to produce soap. John's grandmother made soap by boiling hog fat with wood ashes — which contain potassium and sodium hydroxides. She then skimmed off the soap and pressed it into cakes. Unfortunately, Granny wasn't very good at getting all the proportions just right and tended to use too much base, making the soap very alkaline. In this kind of reaction, called a *saponification reaction,* hydrolysis of the ester groups in the presence of a base yields glycerol and the carboxylate ions of the three fatty acids. A soap is really a sodium or potassium salt of a fatty acid. The calcium and magnesium analogues, on the other hand, are insoluble. If the soap is used with *hard* water (containing calcium or magnesium ions), it precipitates as a greasy scum: bathtub ring.

Acids also catalyze the hydrolysis of a fat to produce glycerol and a fatty acid. Acid hydrolysis is reversible, whereas the presence of excess base inhibits the reverse of saponification. During digestion, lipases break down triglycerides, and bile salts make the fatty acid portions soluble. A *lipase* is an enzyme that catalyses the decomposition of a fat. *Bile salts* are oxidation products of cholesterol that act as detergents to make the products of the breakdown soluble. In humans, absorption of the products occurs in the small intestine.

No Simpletons Here: Complex Lipids

So far, we have been discussing simple lipids. However, some lipids are somewhat more complex. In general, complex lipids are esters of glycerol or some other alcohol. The two major categories of complex lipids are the phosphoglycerides and the sphingolipids. The *phosphoglycerides* are the plasmalogens and the phosphatidates. The *sphingolipids* are the glycosphingolipids and the sphingophospholipids. (Further subdivision is shown back in Figure 8-1.) A phospholipid is either a phosphoglyceride or a sphingophospholipid. Phospholipids are major components of membranes. Any carbohydrate-containing lipid is a glycolipid. The classifications of lipids overlap. (As you may have noticed, nothing in biochemistry is ever truly simple.) For this reason, a lipid may fall into more than one subcategory.

Phosphoglycerides

The alcohol here is glycerol, to which two fatty acids and a phosphoric acid are attached as esters. This basic structure is a phosphatidate. *Phosphatidate* is an important intermediate in the synthesis of many phosphoglycerides. The presence of an additional group attached to the phosphate allows for many different phosphoglycerides.

By convention, structures of these compounds show the three glycerol carbon atoms vertically with the phosphate attached to carbon atom number three (at the bottom). The occurrence of phosphoglycerides is almost exclusive to plant and animal cell membranes. Plasmalogens and phosphatidates are examples. These are also known as glycerophospholipids.

Plasmalogens

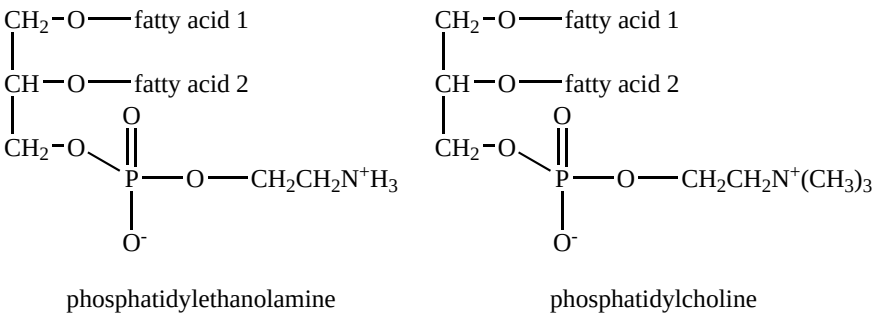
Plasmalogens are a type of phosphoglyceride. The first carbon of glycerol has a hydrocarbon chain attached via an ether, not ester, linkage. Ether linkages are more resistant to chemical attack than ester linkages are. The second (central) carbon atom has a fatty acid linked by an ester. The third carbon links to an ethanolamine or choline by means of a phosphate ester. These compounds are key components of the membranes of muscles and nerves.

Phosphatidates

Phosphatidates are lipids in which the first two carbon atoms of the glycerol are fatty acid esters, and the third is a phosphate ester. The phosphate serves as a link to another alcohol — usually ethanolamine, choline, serine, or a carbohydrate. The identity of the alcohol determines the subcategory of the phosphatidate. There is a negative charge on the phosphate and, in the case of choline or serine, a positive quaternary ammonium ion. (Serine also has a negative carboxylate group.) The presence of charges gives a “head” with an

overall charge. The phosphate ester portion ("head") is hydrophilic, whereas the remainder of the molecule, the fatty acid "tail", is hydrophobic. These are important components for the formation of lipid bilayers.

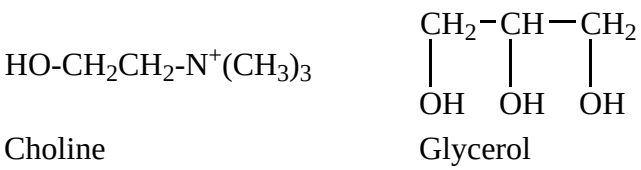
Phosphatidylethanolamines, phosphatidylcholines, and other phospholipids are examples of phosphatidates. Figure 8-5 illustrates examples of a phosphatidylethanolamine and a phosphatidylcholine.



phosphatidylethanolamine phosphatidylcholine

Figure 8-5: Examples of the general formulas of a phosphatidylethanolamine and a phosphatidylcholine.

The structures of some of the alcohols present in lipids appear in Figure 8-6.



Choline Glycerol

Ethanolamine Serine

Figure 8-6: Alcohol components of lipids.

Phosphatidylethanolamines

These are the most common phosphoglycerides in animals and plants. In animals, many of these are the *cephalins,* which are present in nerves and brain tissue. They are also factors involved in blood clotting. Recall that the phosphate has a negative charge and that the nitrogen of the enthanolamine is a quaternary ammonium ion with a positive charge.

Phosphatidylcholines

These are the lecithins. Choline is the alcohol, with a positively charged quaternary ammonium, bound to the phosphate, with a negative charge. Lecithins are present in all living organisms. An egg yolk has a high concentration of lecithins — which are commercially important as an emulsifying agent in products such as mayonnaise. Lecithins are also present in brain and nerve tissue.

Other phospholipids

There are many other phospholipids, some of which are glycolipids. The glycolipids include phosphatidyl sugars where the alcohol functional group is part of a carbohydrate. Phosphatidyl sugars are present in plants and certain microorganisms. A carbohydrate is very hydrophilic due to the large number of hydroxyl groups present.

Sphingolipids

Sphingolipids occur in plants and animals, and are especially abundant in brain and nerve tissue. In these lipids, sphingosine (Figure 8-7) replaces glycerol. The alcohol groups in the sphingosine may form esters just like the similar groups on glycerol. The amino group can form an amide. The combination of a fatty acid and sphingosine, via an amide linkage, is a *ceramide,* which is an intermediate in the formation of other sphingolipids.

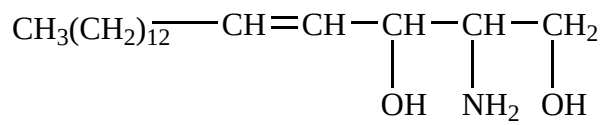


Figure 8-7: Structure of sphingosine.

Glycosphingolipids

A *glycosphingolipid* is an important membrane lipid containing a carbohydrate attached to a ceramide. The carbohydrate serves as a polar (hydrophilic) head. The carbohydrate may be either a monosaccharide or an oligosaccharide. The carbohydrate sequence in the oligosaccharide is important in helping these compounds recognize other compounds in biochemical reactions sequences. The carbohydrate portion is always on the outside of the membrane.

Cerebrosides

A *cerebroside* consists of a monosaccharide attached to a ceramide. The carbohydrate is either glucose or galactose. Cerebrosides are present in nerve and brain cells, though most animal cells contain some of these compounds.

Gangliosides

Gangliosides are sphingolipids with complex structures. The ceramide has an oligosaccharide, containing three to eight monosaccharide units, attached. The monosaccharide units may or may not be substituted. They are very common as part of the outer membranes of nerve cells, where the sugar sequence leads to cell recognition and communication. Small quantities of gangliosides are part of the outer membranes of other cells. When present in a membrane, the carbohydrate portion is always extracellular.

Sphingophospholipids

Sphingophospholipids contain sphingosine, a fatty acid, phosphate, and choline. An example is sphingomyelin, which is an important constituent of the myelin sheath surrounding the axon of all nerve cells. Multiple sclerosis, among other diseases, is a consequence of a fault with the myelin sheath. Sphingomyelin is the most common of the sphingolipids, and it is the only sphingosine phospholipid found in membranes.

Membranes: The Bipolar and the Bilayer

One use of lipids is in the construction of membranes. *Membranes* are used to separate regions both in and around cells — a typical membrane, as shown in Figure 8-8, is a lipid bilayer or bimolecular sheet. The polar portions of the lipids, the heads, are on the outside edges of the bilayer, whereas the nonpolar portions, the tails, are in the interior. The heads of the lipids appear as circles in our illustrations, and the tails appear as strings. The tails are usually long fatty acid chains. The hydrophilic heads, often with a charge, are in contact with aqueous material, and the hydrophobic tails are away from the aqueous material. Interactions between the hydrophobic tails are the key factors leading to the formation of lipid bilayers. *Lipid bilayers* tend to form closed structures or compartments to avoid having exposed hydrophobic edges. The membranes tend to be self-sealing.

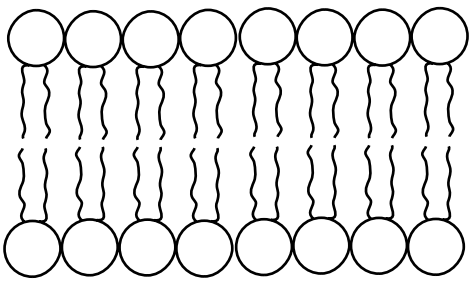


Figure 8-8: A simplified representation of a lipid bilayer.

Actual cell membranes are not as symmetrical as the one shown in Figure 8-8. This asymmetry is due in part to the presence of other components, and in part to differences between the intracellular and extracellular surfaces. If the fatty acid portions are not saturated, the tails will not form parallel structures, and there will be "holes" present within the bilayer. These holes are an essential feature leading to membrane fluidity. Other components include proteins and cholesterol. The carbohydrate portion of glycolipids is on the extracellular side of the bilayer instead of the intracellular side.

Polar materials cannot readily pass through the hydrophobic region of membranes, and nonpolar materials cannot readily pass through the hydrophilic outer region. Water, due to its small size and high concentration, can transverse the bilayer faster than ions and most other polar molecules. In actual cells, certain mechanisms allow material to cross the bilayer but require other components to be present in the bilayer. These components, mostly proteins, give selective permeability of the membranes. In addition, other materials, such as cholesterol, are necessary to serve other functions, such as stiffening the membrane.

Membranes may contain roughly from 20 to 80 percent protein, which may be *peripheral* (on the surface of the membrane) or *integral* (extending into or through the membrane). Integral proteins interact extensively with the hydrophobic portion of the bilayer, as illustrated in Figures 8-9 and 8-10.

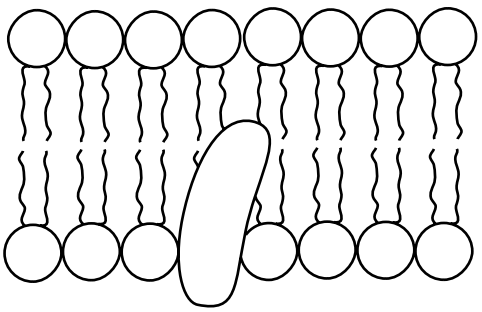


Figure 8-9: An integral protein that does not pass through the membrane.

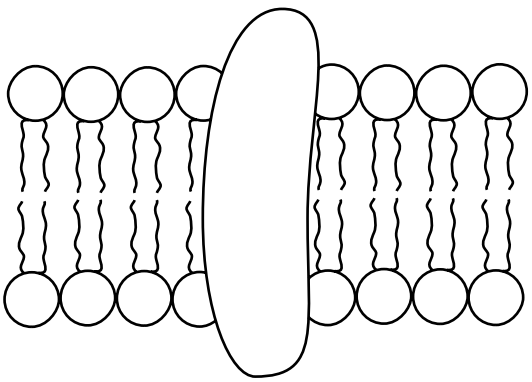


Figure 8-10: An integral protein passing through the membrane.

Peripheral proteins typically bind to the surface through electrostatic or hydrogen bonding, although covalent interactions are possible. Proteins are important for most membrane processes. If the protein is a glycoprotein, the carbohydrate portion lies on the external side of the membrane and is important to intercellular recognition.

Crossing the wall: Membrane transport

A lipid bilayer is, by its nature, impermeable to polar molecules and ions (hydrophilic species). Nevertheless, cells need to be able to bypass this feature and get hydrophilic materials in and out. There are two ways to circumvent impermeability: A *pump* involves active transport using energy to work against a concentration gradient, and a *channel* involves passive transport or facilitated diffusion using a concentration gradient.

Nonpolar molecules are lipophilic and dissolve in the lipid bilayer. In general, lipophilic materials pass through the membrane by simple diffusion along a concentration gradient. Channels and pumps are mainly to allow hydrophilic species to transverse the hydrophobic region of the bilayer.

Pumps

Pumps require energy to function. In many cases, the hydrolysis of ATP provides the needed energy. The generic name for this type of pump is a P-type ATPase. The name derives from the transfer of a phosphate from an ATP to an intermediate, a step that is essential to the action of the pump. Pumps can transfer other species than ions.

When is a solid a liquid? The fluid mosaic model

The lipid bilayer structure gives much insight into the structure of membranes but little information about their function. Many functions of the membrane depend upon its fluidity, best described by using the *fluid mosaic model.* In this model, the membrane serves as a permeability barrier and as a solvent for the integral proteins. Diffusion along the plane of the membrane — *lateral diffusion* — of the membrane components is often rapid. In general, lipids move more rapidly than proteins, with some proteins being essentially immobile. Diffusion of membrane components across the membrane — *transverse diffusion* — is usually slow.

The fluidity of the membrane depends on a number of factors. Bacteria adjust the fluidity by utilizing fatty acid chains — longer chains are less fluid than are shorter chains. The presence of double bonds makes the membrane more fluid. In animals, cholesterol controls the fluidity: The greater the cholesterol concentration, the less fluid the membrane. The transition from the rigid to the fluid state occurs at a temperature known as the *melting temperature*, T_m.

Most animal cells have a high potassium ion and a low sodium ion concentration relative to the extracellular environment. It requires energy to generate and maintain this gradient. The transport system is the Na^+-K^+ pump, also referred to as Na^+-K^+ ATPase. Hydrolysis of ATP provides the energy to transport potassium ions into the cell and sodium ions out of the cell. Both the sodium and potassium ions must be simultaneously bound to the pump. The pump simultaneously transports three sodium ions out of the cell as it transports two potassium ions in.

Not all pumps require the hydrolysis of ATP to supply energy. Some utilize the transport on one species to facilitate the transport of another. The transport of one species with the concentration gradient can pump another against the concentration gradient. The responsible membrane proteins are the cotransporters or secondary transporters. *Cotransporters* may be either symporters or antiporters. In a symporter, both transported species move in the same direction, whereas in an antiporter, the species move in opposite directions. The sodium-calcium exchanger is an example of an antiporter, which pumps three sodium ions into a cell for every two calcium ions pumped out. Some animal cells use a symporter to pump glucose coupled with sodium ions into the cells.

Channels

A channel provides a means of passively transporting a species across a membrane. It is possible to transport a species through a channel more than 1,000 times as fast as a pump's. A channel is technically a tube running through the membrane, but its behavior is significantly more complicated.

Channels are highly selective. Some select on size — sodium is smaller than potassium — whereas others differentiate between anions and cations. A channel exists in an open state to allow transport and a closed state to inhibit it. Some type of regulation is required to convert a channel between an open and a closed state. When a chemical potential regulates the channel, it is a voltage-regulated gate. The regulation may be due to specific chemicals. Chemically controlled regulation is ligand-gated. After the appropriate regulator is removed, the open channels will spontaneously close.

The best-known ligand-gated channel is the acetylcholine receptor. This channel is important for the transmission of nerve impulses. When a nerve impulse reaches the junction between one nerve and the next — the synapse — it triggers the release of acetylcholine, which transverses the small gap to the next nerve and attaches to acetylcholine receptors. This attachment opens the channel, leading to inward sodium ion diffusion and outward potassium ion diffusion. The change in the ion concentrations transmits the nerve impulse into the second nerve cell.

The increase in the sodium ion concentration in the second nerve cell triggers a mechanism to remove sodium ions from the nerve cell. Later another gate brings potassium ions back into the cell.

Steroids: Pumping up

Steroids are another class of lipids. All steroids have the basic core shown in Figure 8-11. A, B, C, and D are common labels for the rings. Different steroids have additions to this basic structure; these may include side chains, other functional groups, and unsaturation or aromaticity of the rings.

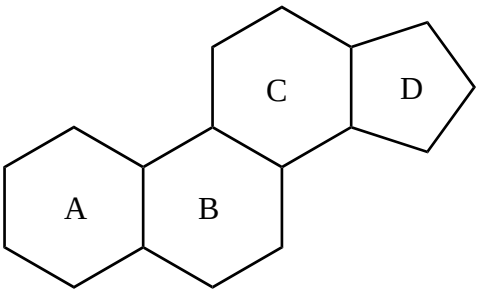


Figure 8-11: Basic structure of a steriod.

Cholesterol is the most abundant steroid. It is a membrane component and serves as the source of other steroids and related materials. Cholesterol comes from the diet, but if insufficient cholesterol is available there, it is synthesized in the liver. The steroid hormones are regulators produced from cholesterol.

Bile salts (mentioned earlier) are a group of materials produced by the oxidation of cholesterol. Unlike cholesterol and the other lipids, bile salts are soluble in water. They are useful as "detergents" to aid in digestion.

The steroids you hear about in the news being used by athletes and bodybuilders are *anabolic steroids,* which increase the body's ability to prevent muscle breakdown and to actually increase the ability to produce muscle. They have structures similar to testosterone, whose function is to enhance male characteristics such as facial hair and muscle mass. However, steroids in large doses have serious side effects: impotence, reduced testicle size, liver tumors, enlargement of the heart, enlargement of the breasts in men, aggressive behavior, and so on. (Sounds great, doesn't it?) Their use without a valid prescription has been illegal since 1991.

Prostaglandins, Thromboxanes, and Leukotrienes: Mopping Up

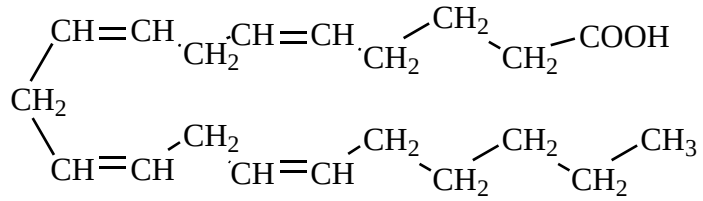
Arachidonic acid, a 20-carbon, polyunsaturated fatty acid, serves as the direct or indirect starting material for the formation of prostaglandins, thromboxanes, and leukotrienes. Cells synthesize both leukotrienes and prostaglandins from arachidonic acid. Additional prostaglandins and thromboxanes come from the prostaglandin derived from arachidonic acid. All three classes of compounds are local hormones. Unlike other hormones, they are not transported via the bloodstream. They are short-lived molecules that alter the activity of the cell producing them and neighboring cells.

All of these compounds are extremely potent chemicals that serve as hormone mediators. They also have many other medical applications and can cause medical problems. They are also known as *eicosanoids* — from the Greek for *twenty,* which alludes to the presence of 20 carbon atoms (Figure 8-12).

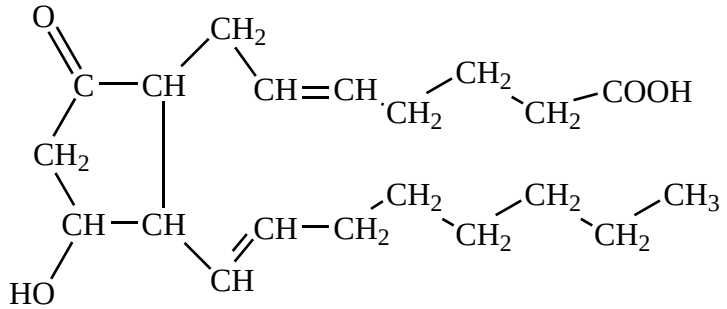
The name *prostaglandin* came from the belief that the prostate gland was its source because they were first isolated from seminal fluid in 1935. Now we know that they are produced in a very wide variety of cells. Prostaglandins differ slightly from each other, but they all contain a five-carbon ring. These minor difference lead to distinct behaviors, although all prostaglandins lower blood pressure, induce contractions in smooth muscles, and are part of the inflammatory response system.

A number of medications are synthetic prostaglandins. For example, derivatives of the prostaglandin PGE_2 are useful in inducing labor. Prostaglandins associated with inflammation are the main cause of the associated redness, pain, and swelling. The half-life of many prostaglandins is only a few minutes or less. Platelets in the blood generate thromboxanes to serve as vasoconstrictors and to induce aggregation of the platelets, two steps leading to the formation of a blood clot. Thromboxane A_2 is an example of one of these agents that induces blood clotting. White blood cells, leukocytes, and other tissues produce leukotrienes, whose name refers to where they were first discovered (leukocytes) and to the presence of three conjugated double bonds (triene). Leukotrienes are associated with allergy attacks.

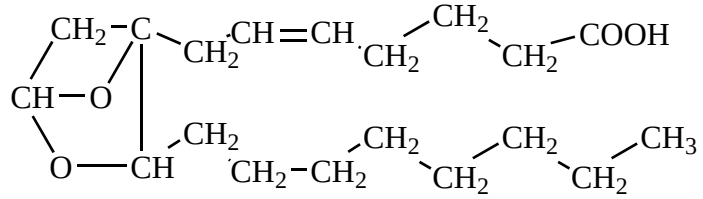
Aspirin interferes with the synthesis of prostaglandins and thromboxanes. Aspirin is an anti-inflammatory agent because it counters the inflammation induced by prostaglandins. The interference with the formation of thromboxanes may be part of the reason why low doses of aspirin help prevent heart attacks and strokes. Low thromboxane levels would inhibit blood clotting. Another anti-inflammatory drug, cortisone, inhibits the release of arachidonic acid from cell membranes, which, in turn, inhibits the formation of the eicosanoids. The fatty acids in fish oils inhibit the formation of the more potent leukotrienes and thromboxanes.



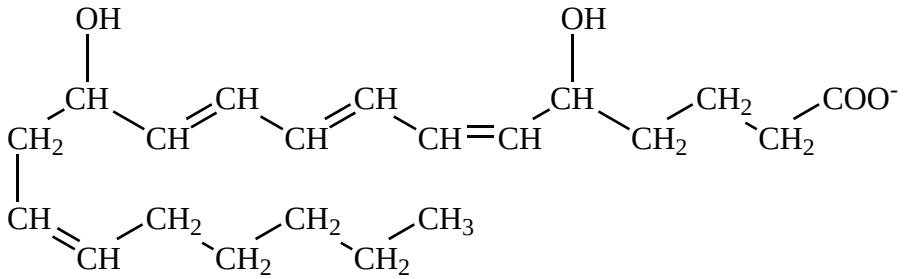
arachidonic acid



a prostaglandin (PGE_2)



a thromboxane (A_2)



a leukotriene (B_4)

Figure 8-12: Structures of arachidonic acid, a typical prostaglandin, thromboxane, and leukotriene.

Chapter 9

Nucleic Acids and the Code of Life

In This Chapter

- Finding out about the structure of proteins
- Understanding amino acid sequencing in proteins
- Going over applications of protein sequencing

Nucleic acids get their name because they were first found in the nuclei of cells. DNA (*deoxyribonucleic acid*) — the most famous nucleic acid — is part of the *chromosomes,* which contain the genes. And the *genes* are ultimately responsible for the synthesis of proteins. Most, if not all, of these proteins are enzymes, each catalyzing a specific chemical reaction occurring in the organism. Indeed, there is a one-gene-one-enzyme hypothesis, where each gene is responsible for the synthesis of one enzyme.

DNA has two direct purposes: It must generate new DNA (replication) so that new generations of cells will have the information necessary to their survival. And it must generate RNA (*ribonucleic acid*). The RNA is involved in the direct synthesis of proteins, called *translation.* These proteins are essential for the maintenance of life.

Nucleotides: The Guts of DNA and RNA

Both DNA and RNA are polymers of nucleotides. A *nucleotide* is a combination of a nitrogen base, a 5-carbon sugar, and a phosphoric acid. There are five different bases present in a nucleotide, and two different sugars. We take a closer look at the components of these nucleotides and then show you how they all fit together.

Reservoir of genetic info: Nitrogen bases

The bases fall into two categories, the general defining structures of which appear in Figure 9-1.

- The purines (adenine and guanine), composed of two fused rings incorporating two nitrogen atoms in each ring and
- The pyramidines (cytosine, thymine, and uracil), composed of a single ring with two nitrogen atoms in the ring structure

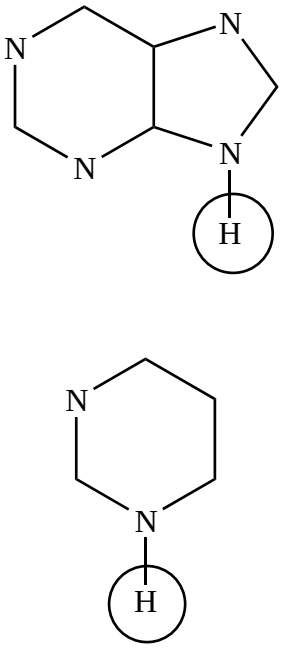
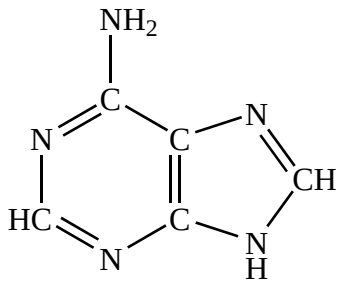


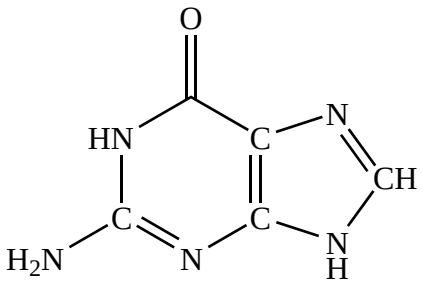
Figure 9-1: Basic purine structure (top) and basic pyramidine structure (bottom).

Adenine (A), guanine (G), and cytosine (C) occur in both DNA and RNA. Thymine (T) is only found in DNA, whereas uracil (U) only occurs in RNA. There are modified forms of some of these bases present in some nucleic acid molecules. The circled hydrogen atoms shown in Figure 9-1 are lost when combining with other components to produce a nucleic acid. The complete structures of the five bases are shown in Figure 9-2. It is the sequence of these bases that stores the genetic information.

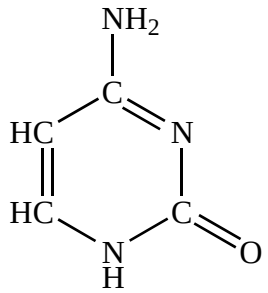
The nitrogen and oxygen atoms present on the nitrogen bases provide a number of sites where hydrogen bonding is possible. Hydrogen bonding is most effective and easily formed between certain combinations of nitrogen bases. Because of this, certain combinations will form, and it is this pattern that is responsible for the transmission of information. The atoms on the nitrogen bases normally use a regular numbering system, whereas the atoms in the sugar component use primed numbers.



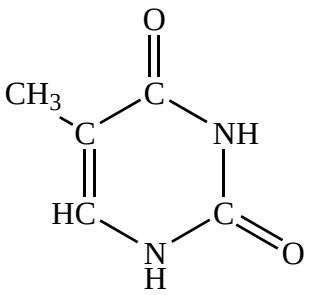
Adenine (A)



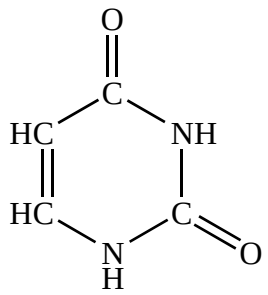
Guanine (G)



Cytosine (C)



Thymine (T)



Uracil (U)

Figure 9-2: Adenine (A), guanine (G), cytosine (C), thymine (T), uracil (U).

The sweet side of life: The sugars

The 5-carbon sugars found in the nucleic acids are D-ribose and D-deoxyribose. The difference between these two sugars is that deoxyribose is missing an oxygen atom on carbon atom number 2'. The structures for these two sugars appear in Figure 9-3. The arrows in the figure point to the alcohol group on carbon atom number 1', the *anomeric* carbon. This is where the nitrogen base will attach. Both sugars adopt the β form of the furanose ring. Numbering of the sugar begins with the anomeric carbon (1') and proceeds clockwise with the $-CH_2OH$ carbon being 5'.

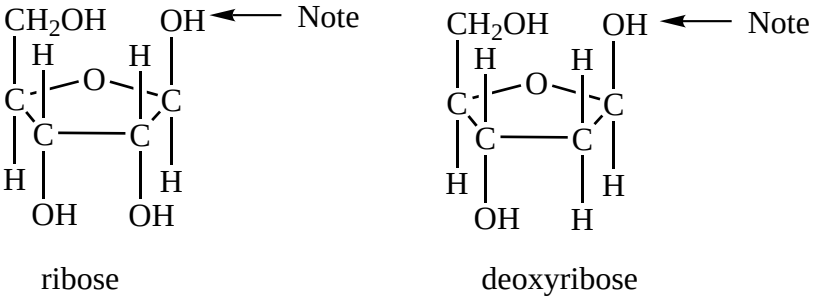


Figure 9-3: Structures of the 5-carbon sugars present in nucleic acids.

The sour side of life: Phosphoric acid

The third component of a nucleotide is a phosphoric acid (Figure 9-4). At physiological pH it does not exist in the fully protonated form shown in the figure. It is responsible for the "acid" in nucleic acid.

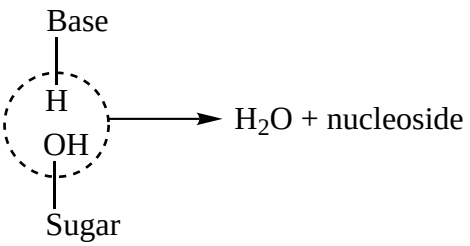


Figure 9-4: Structure of phosphoric acid.

Tracing the Process: From Nucleoside to Nucleotide to Nucleic Acid

Remember Legos and Tinker Toys? Putting together the pieces to get something new? That's what goes on in the construction of nucleic acids. Nature first joins a nitrogen base and a 5-carbon sugar to form a nucleoside; then that nucleoside joins with phosphoric acid to form a nucleotide; finally, the combination of these nucleotides produces a nucleic acid.

First reaction: Nitrogen base + 5-carbon sugar = nucleoside

The combination of a nitrogen base with a 5-carbon sugar is a *nucleoside.* The general reaction appears in Figure 9-5. It is a *condensation* reaction. Remember the condensation reactions you studied in ester formation in organic chemistry? This is exactly the same type. Here a compound containing hydrogen (the nitrogen base) approaches another molecule containing an –OH group (a sugar). The hydrogen combines with the –OH to form water, which is expelled. A bond forms in the remaining fragments.

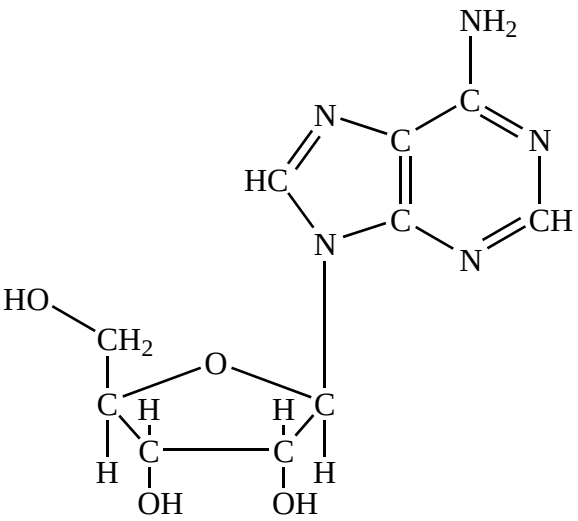


Figure 9-5: General reaction for the formation of a nucleoside.

The name of the nucleoside comes from the nitrogen base if the sugar is ribose; it has a prefix if the sugar is deoxyribose. For example, adenine combines with ribose to form adenosine and combines with deoxyribose to form deoxyadenosine. The structure for the nucleoside adenosine is in Figure 9-6. The hydrogen atom lost from the base was the one circled in Figure 9-1.

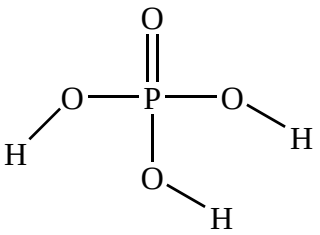


Figure 9-6: Structure of the nucleoside adenosine.

Second reaction: Phosphoric acid + nucleoside = nucleotide

The combination of a phosphoric acid with a nucleoside produces a nucleotide, which is a phosphate ester, as seen in Chapter 3, of a nucleoside. The formation involves a condensation reaction between the phosphoric acid and the alcohol group on carbon number 5, the $-CH_2OH$ (Figure 9-7).

Adenosine monophosphate (AMP) is an example of a nucleotide (Figure 9-8). Nucleotides are the monomers from which nucleic acids form. AMP is not only one of the "Legos" that makes RNA but is also very much involved in the energy transfer process in the cells (much more on AMP in Part IV).

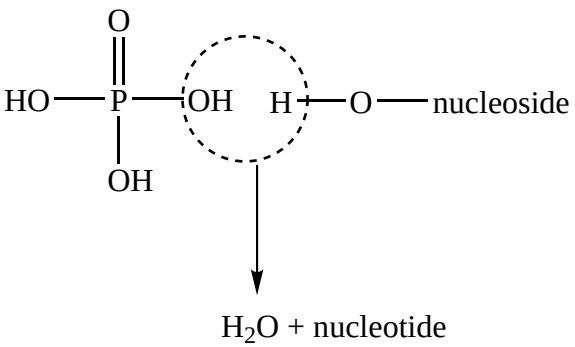


Figure 9-7: Simplified representation of the formation of a nucleotide.

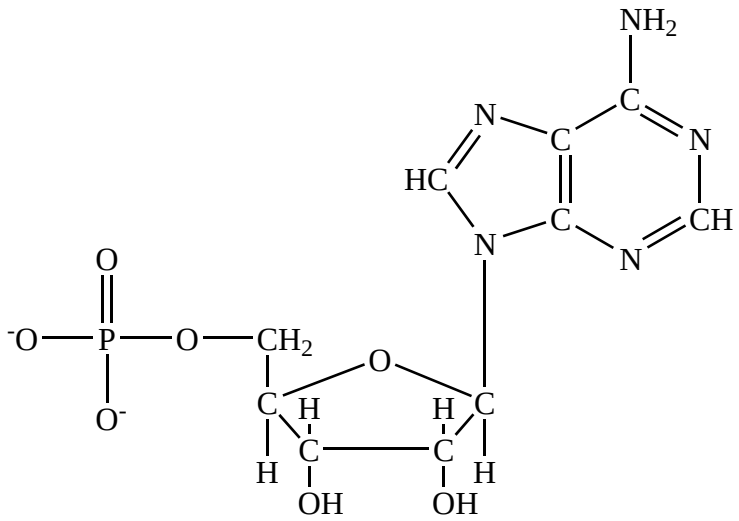


Figure 9-8: Structure of adenosine monophosphate (AMP).

If the sugar is ribose, then the result is one of four ribonucleotides. If the sugar is deoxyribose, the result is one of the four deoxyribonucleotides.

Third reaction: Nucleotide becomes nucleic acid

Nucleic acids form by joining nucleotides using the same condensation reactions we've mentioned. This condensation reaction involves the phosphate of one nucleotide reacting with the alcohol group on carbon atom number 3' of another nucleotide. Figure 9-9 illustrates. Note that the lower –OH, in the circle, is from the phosphoric acid, attached to carbon-5'. The upper –H in the circle is from the alcohol on carbon-3'.

The starting end of the polymer is 5', whereas the terminal end is 3'. Figure 9-10 illustrates the 5' and 3' carbon atoms on adenosine monophosphate.

A Primer on Nucleic Acids

Nucleic acids are responsible for storing and directing the information our cells use for reproduction and growth. They are large molecules found in the cell's nucleus. The genetic information is contained in the DNA, in terms of its primary and secondary structure. As a cell divides and reproduces, the genetic information in the cell is *replicated* to the new cells, which must be done accurately and precisely — no mistakes must be made. RNA's role is to

transfer the genetic information found in the DNA to the ribosomes, where protein synthesis occurs. DNA and RNA allow us to live and function.

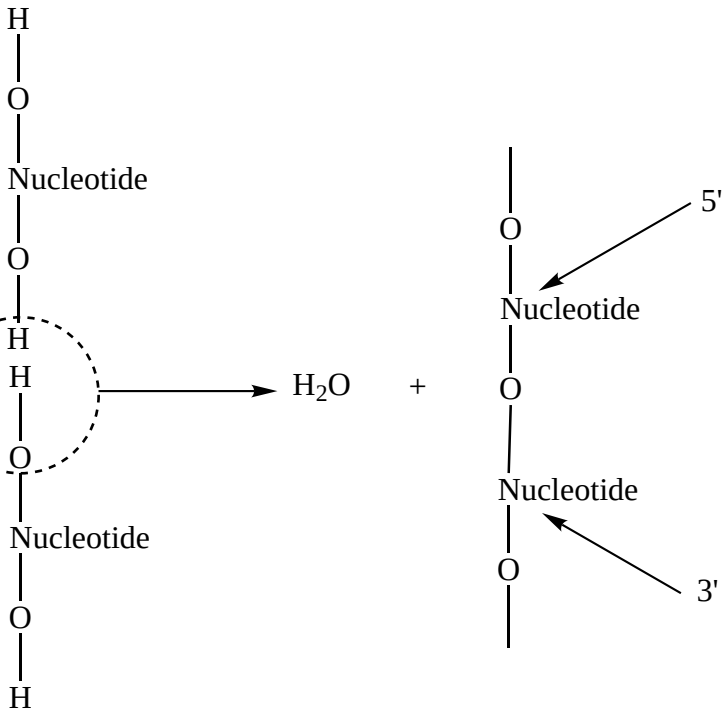


Figure 9-9: Simplified representation of the joining of two nucleotides.

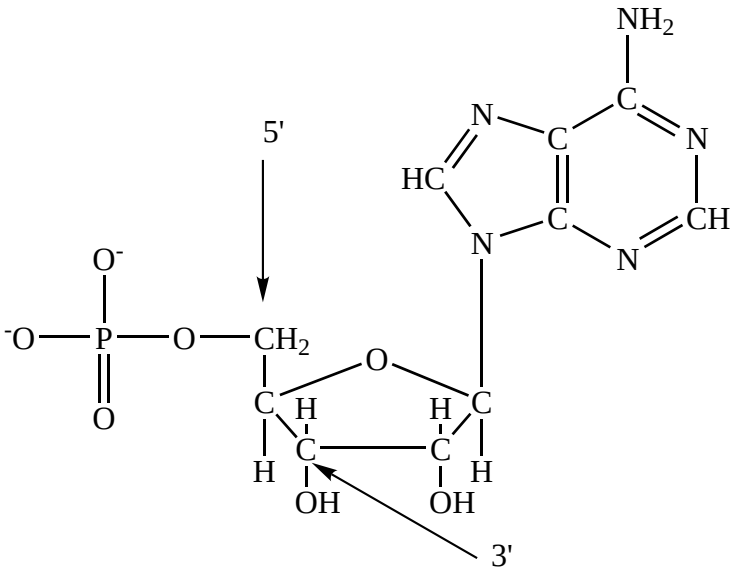


Figure 9-10: 5' and 3' carbon atoms on adenosine monophosphate.

DNA and RNA in the grand scheme of life

Both DNA and RNA are polymers composed of nucleotide subunits. However, DNA is a much larger molecule than RNA. DNA molecules typically have molecular weights in the billions. The human genome contains about 3 billion nucleotides.

As a simplification, the structure of a particular nucleic acid may be represented as 5'-C-G-T-A-3'. This abbreviation indicates that we begin at the 5' end and end at the 3' end (as always), and the nitrogen bases on the nucleotides are, in order, cytosine (C), guanine (G), thymine (T) and adenine (A). There are three different types of RNA, and each one has a specific use:

- Ribosomal RNA (rRNA) is the most common: 75–80 percent occurs within the ribosomes of the cell.
- Transfer RNA (tRNA) accounts for 10–15 percent.
- Messenger RNA (mRNA) makes up the remainder.

All three types are important to protein synthesis — which occurs in the ribosomes, home of ribosomal RNA (rRNA). The amino acids necessary for protein synthesis are transferred to the ribosomes by transfer RNA (tRNA). The message instructing the ribosomes how to assemble the protein travels from the DNA to the ribosome via messenger RNA (mRNA). This message tells the ribosome the sequence of amino acids to make a specific protein.

Transfer RNA contains the fewest nucleotides: 70–90. The average mRNA has about 1,200 nucleotides. There are three subcategories of rRNA ranging from about 120 to over 3,700 nucleotides. (DNA typically has between 1 million and 100 million nucleotides, though viral DNA tends to be smaller.) *Ribonucleotides* have other uses in addition to building RNA. They are present in energy molecules (ATP), in intracellular hormone mediator (cyclic AMP), and in certain coenzymes (FAD and NAD^+). Plants and animals contain both DNA and RNA. Viruses can contain either DNA or RNA.

Nucleic acid structure

The structure of a particular nucleic acid controls its function within the organism. For example, the structure of a particular tRNA determines which specific amino acid it will transfer to the ribosome for protein synthesis. In fact, the difference between DNA and RNA resides in the structure of the molecules. Because of the complexity of these types of molecules, there may be more than one key type of structure present.

The primary structure of the nucleic acids is the sequence of nucleotides, the order in which the individual nucleotides have been joined. This sequencing determines which hydrogen bonds form, and this, in turn, controls much of the function of the nucleic acid. DNA also has an important secondary structure, a consequence of hydrogen bonding between the nitrogen bases on the DNA strands. The result is that DNA consists of a *double helix* — which looks like a ladder twisted lengthwise — where hydrogen bonds (the rungs in the ladder) hold the two primary structures together.

The hydrogen bonds between the two stands of DNA make the two strands *complementary (*paired). Every A is complementary to a T, and every G is complementary to a C in *base pairing.* Base pairing is essential for the function of the nucleic acids.

The two DNA strands are *antiparallel,* which means that the 5' end of one strand connects to the 3' end of its complementary strand. This pairing also places the more polar (more hydrophilic) sugar and phosphate groups on the outside and the less polar (more hydrophobic) nitrogen bases on the inside. (Note that *hydrophilic* and *hydrophobic* as used here are relative terms.) The antiparallel nature affects how DNA produces new DNA (the replication process) and new RNA (the transcription process).

Although each of the nitrogen bases is very efficient at forming hydrogen bonds, certain combinations are extremely effective. In DNA, an adenine is capable of forming two hydrogen bonds to thymine (Figure 9-11), and guanine can form three hydrogen bonds to cytosine (Figure 9-12).

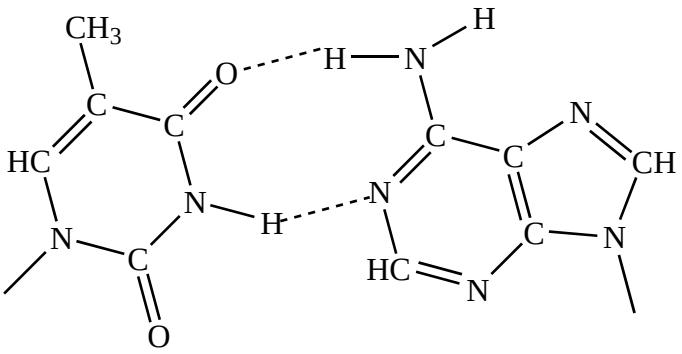


Figure 9-11: Hydrogen bonds (dotted lines) form between adenine (right) and thymine (left).

Adenine is also able to form hydrogen bonds with uracil when DNA interacts with RNA or when two RNA molecules interact. The interaction between adenine and uracil is shown in Figure 9-13.

The ability to form these specific combinations is important in real life — this is the *genetic code* we all have heard so much about. The sequencing of nucleotides in the nucleic acids and the sequencing of amino acids in the proteins all depend on these hydrogen bonds. Without them, the appropriate information would not be transferred precisely, and you might produce kittens instead of kids. The result? DNA, the structure of life (Figure 9-14).

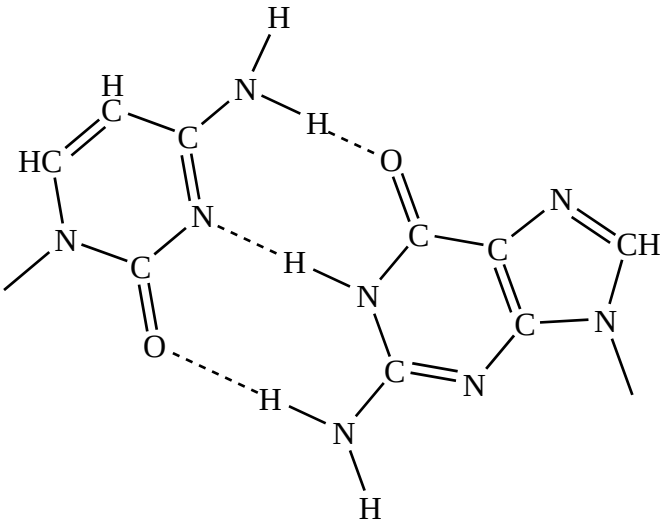


Figure 9-12: Hydrogen bonds (dotted lines) form between guanine (right) and cytosine (left).

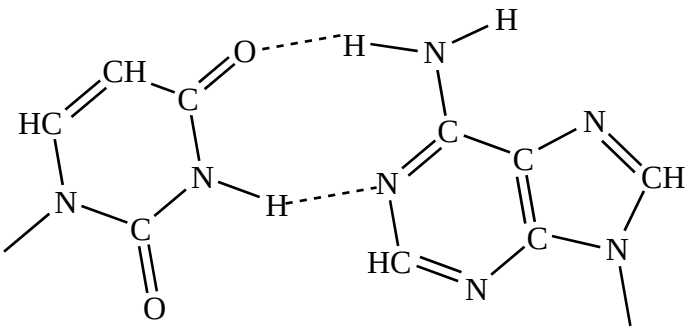


Figure 9-13: Hydrogen bonds (dotted lines) form between adenine (right) and uracil (left).

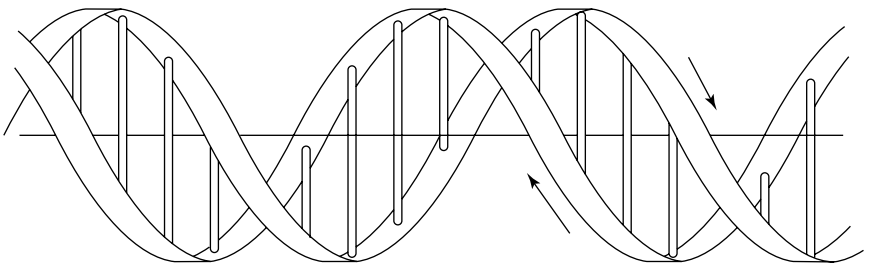


Figure 9-14: The secondary structure of DNA.

Chapter 10

Vitamins and Nutrients

In This Chapter

- Taking a look at the purpose of vitamins
- Understanding B vitamins
- Assessing other vitamins and nutrients

An organism must absorb a variety of materials to live, many of which fall into the category of food, certainly one of *our* favorite categories, especially John's. These foodstuffs required by an organism for life and growth are classified as nutrients. *Nutrients* are the substances in the diet necessary for growth, replacement, and energy. Here are the six general classes of nutrients:

- Carbohydrates
- Lipids
- Proteins
- Vitamins
- Minerals
- Water

Digestion converts large molecules in food into smaller molecules that can be absorbed. During digestion, carbohydrates (with the exception of the monosaccharides), lipids, and proteins are broken down into their components. These components are often used by the organism directly for growth and replacement. For animals, energy comes primarily from carbohydrates and lipids, but proteins can also serve as an energy source.

Vitamins are other organic materials required by an organism, and minerals are inorganic materials required by an organism. In addition, all living organisms require water to survive. Water is a wonderful substance. For more about the unusual properties of water, check out Chapter 2 in this book or *Chemistry For Dummies* by John T. Moore (Wiley).

More than One-a-Day: Basics of Vitamins

Vitamins are organic compounds that are required, in small quantities, for normal metabolism. The term *active form* is used to describe the structural form of the molecule, in this case vitamins, that performs its function (exhibits *activity*) within the organism. In general, humans cannot synthesize sufficient quantities of vitamins; thus, vitamins must come from other sources — through the diet and/or in pill form. A deficiency of a vitamin in the diet leads to a health problem. The general symptoms for any vitamin deficiency include frequent illness, slow healing of wounds, and tiredness. It was not until the early 1900s that the need for trace nutrients such as vitamins and minerals was first established.

There are two categories of vitamins: water-soluble and fat-soluble. *Water-soluble* vitamins include vitamin C and the B vitamins. Vitamins A, D, E, and K comprise the other category, the *fat-soluble* vitamins. Water-soluble vitamins tend to have more oxygen and nitrogen in their structure than fat-soluble vitamins, which have significant hydrocarbon portions in their structure. The majority of water-soluble vitamins either act as coenzymes or are important in the synthesis of coenzymes. Fat-soluble vitamins serve a variety of biochemical functions.

The body can easily eliminate an excess of the water-soluble vitamins, normally in the urine. The bright yellow of the urine of a person taking large doses of vitamin C attests to that fact. Because the body does not store water-soluble vitamins, continual replacement is necessary. The body can store excess amounts of a fat-soluble vitamin in the body's fatty tissue, and therefore elimination is not very easy. Unfortunately, this can lead to an accumulation of these vitamins, sometimes to toxic levels. One should consider this before consuming mega quantities of the fat-soluble vitamins.

To B or Not to B: B Complex Vitamins

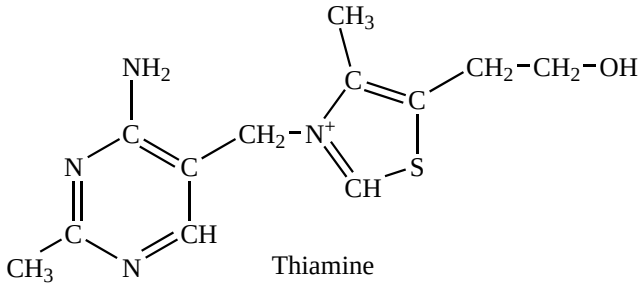
The B vitamins — or B complex — comprise a number of water-soluble vitamins that are found together in a number of sources. Originally, this mixture was thought to be only one vitamin (vitamin B). With the possible exception of vitamin B_6, these appear to be relatively nontoxic. In general, the B complex is important for healthy skin and nervous systems.

Vitamin B_1 (thiamine)

Thiamine is important to carbohydrate metabolism. Like the other B vitamins, the body does not store it. In addition, prolonged cooking of food can destroy it. Once absorbed in the body, thiamine is converted to a form that is biologically active through the attachment of a pyrophosphate (diphosphate) group to give thiamine pyrophosphate (TPP). The structures of vitamin B_1 and thiamine pyrophosphate are shown in Figure 10-1.

TPP is a coenzyme used in decarboxylating pyruvate to acetyl-CoA and α-ketoglutarate to succinyl-CoA. In addition, TPP is necessary for the synthesis of ribose.

A deficiency in thiamine leads to beriberi, which causes deterioration in the nervous system. Beriberi was prevalent in regions where rice was a major food source. Rice, particularly polished rice, is low in thiamine. Using brown rice, which has more thiamine, alleviates this problem. Nursing infants are particularly at risk when their mothers have a thiamine deficiency. Many alcoholics also suffer from this condition because many "foods" high in alcohol are particularly low in vitamins.



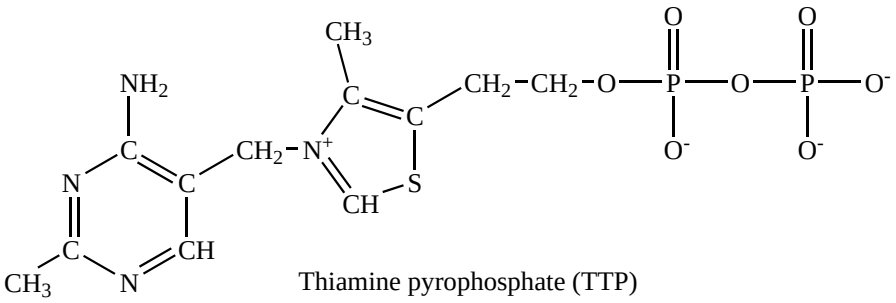


Figure 10-1: Structures of vitamin B^1 (thiamine) and thiamine pyrophosphate (TPP).

Good dietary sources of thiamine include liver, spinach, green peas, navy and pinto beans, whole-grain cereals, and most legumes.

Vitamin B_2 (riboflavin)

Riboflavin is essential to the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The structures of these materials are shown in Figure 10-2. FMN and FAD are important coenzymes involved in a number of biochemical redox processes. The name *riboflavin* alludes to the presence of *ribitol,* an alcohol derived from ribose. The other part of riboflavin is the ring system isoalloazine, a flavin derivative.

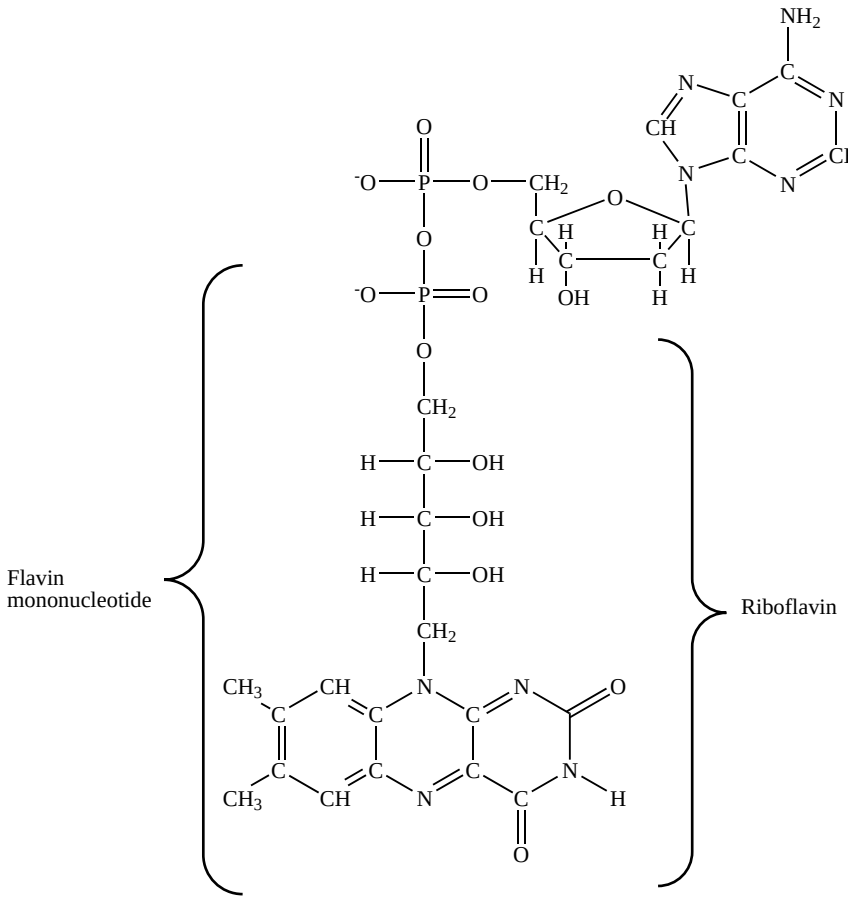


Figure 10-2 : Structure of flavin adenine dinucleotide (the entire structure) and the component materials flavin mononucleotide and riboflavin.

No deficiency diseases are associated with riboflavin; however, a deficiency does lead to burning and itchy eyes, dermatitis, and anemia. Dietary sources of this vitamin include soybeans, liver, milk, cheese and green leafy vegetables. Riboflavin is stable during cooking, but is broken down by light.

Vitamin B_3 (niacin)

The term *niacin* applies to two compounds: nicotinic acid and nicotinamide. These two compounds along with nicotinamide adenine dinucleotide (NAD^+) appear in Figure 10-3. Nicotinamide is part of the coenzymes NAD^+ and nicotinamide dinucleotide phosphate ($NADP^+$). These coenzymes work with a number of enzymes in catalyzing a number of redox processes in the body.

Niacin is one of the few vitamins that the body *can* synthesize. The synthesis utilizes tryptophan and is not very efficient.

Pellagra is a niacin-deficiency disease. Symptoms include loss of appetite, dermatitis, mental disorders, diarrhea, and possibly death. It was common in the southern United States in the early 1900s because many people had a diet of corn, which is neither a good source of niacin or tryptophan.

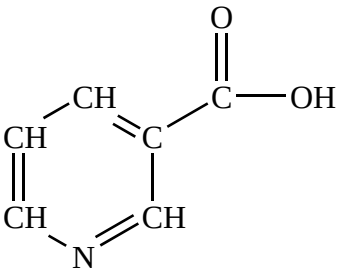
There are many dietary sources for niacin, including most meats and vegetables, milk, cheese, and grains.

Vitamin B_6 (pyridoxine)

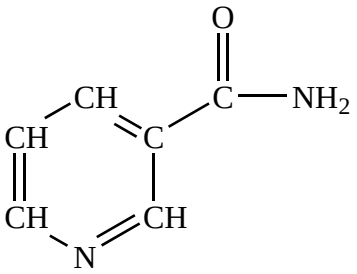
This vitamin consists of three components: pyridoxine, pyridoxal, and pyridoxamine. All three need to be converted to pyridoxal phosphate, a form that is biologically active in the organism. The structures for these compounds appear in Figure 10-4. Pyridoxal phosphate serves as a coenzyme in a variety of processes, including the interconversion of α-keto acids and amino acids.

Avocados, chicken, fish nuts, liver, and bananas are especially good food sources of vitamin B_6. Heating decreases its concentration in food.

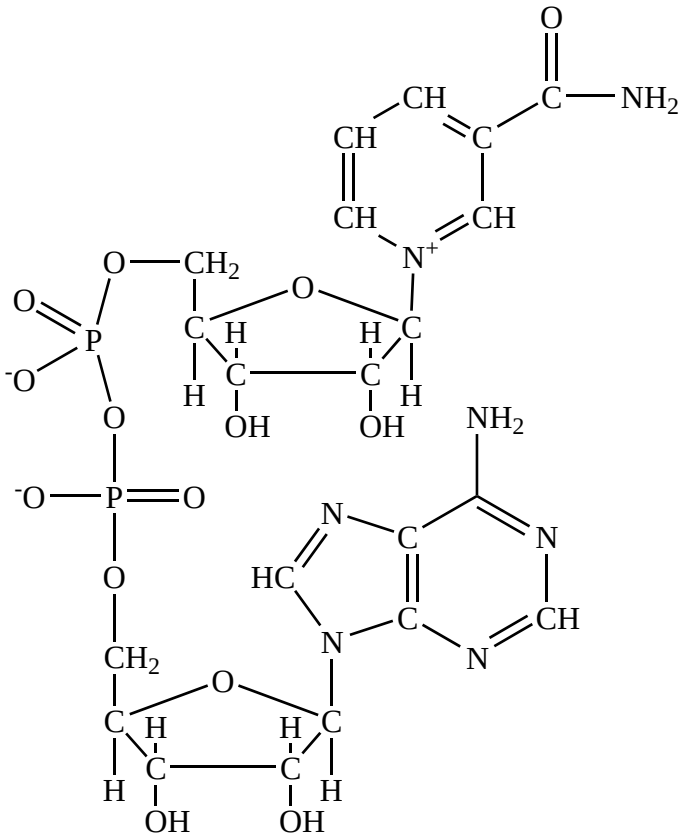
There is no pyridoxine-deficiency disease known; however, low levels can lead to irritability, depression, and confusion. Unlike the other water-soluble vitamins, there is evidence that large doses of vitamin B_6 may lead to health problems. The symptoms of excess vitamin B_6 consumption include irreversible nerve damage.



Nicotinic acid

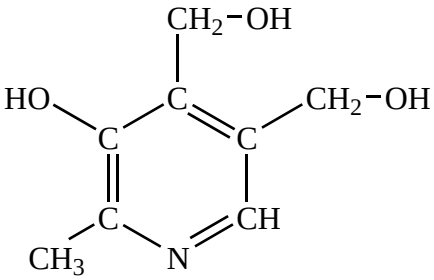


Nicotinamide

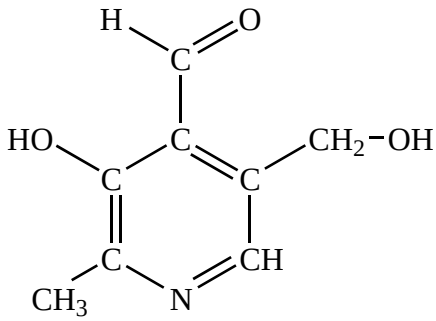


Nicotinamide adenine dinucleotide (NAD^+)

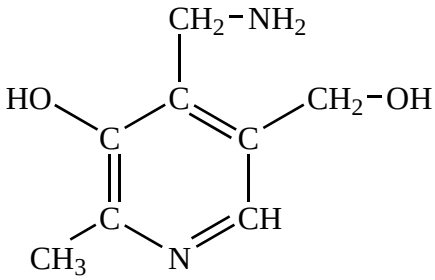
Figure 10-3: Structures of nicotinic acid, nicotinamide, and nicotinamide adenine dinucleotide (NAD^+).



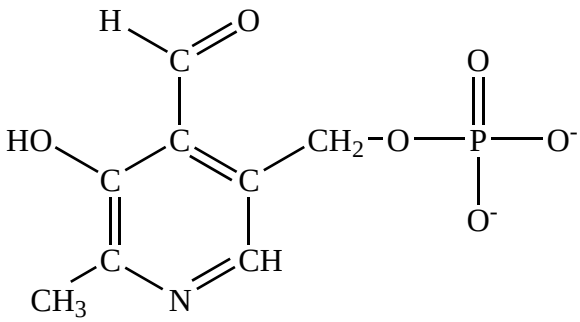
Pyridoxine



Pyridoxal



Pyridoxamine



Pyridoxal phosphate

Figure 10-4: Structures of pyridoxine, pyridoxal, pyridoxamine, and pyridoxal phosphate.

Biotin

Biotin is a coenzyme important to many carboxylation reactions. Biotin is the carbon transporter in both lipid and carbohydrate metabolism.

Bacteria in the intestinal track synthesize biotin in sufficient quantities to minimize the chances for a deficiency. However, antibiotics can inhibit the growth of these bacteria and induce a deficiency. In these circumstances, the symptoms include nausea, dermatitis, depression, and anorexia. Biotin is stable to cooking. Its structure is shown in Figure 10-5.

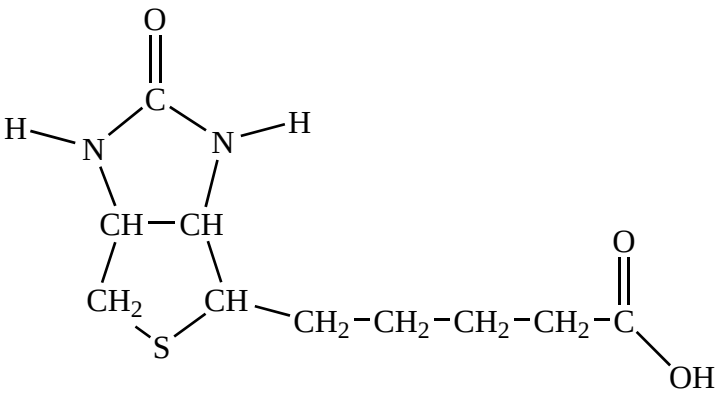


Figure 10-5: Structure of biotin.

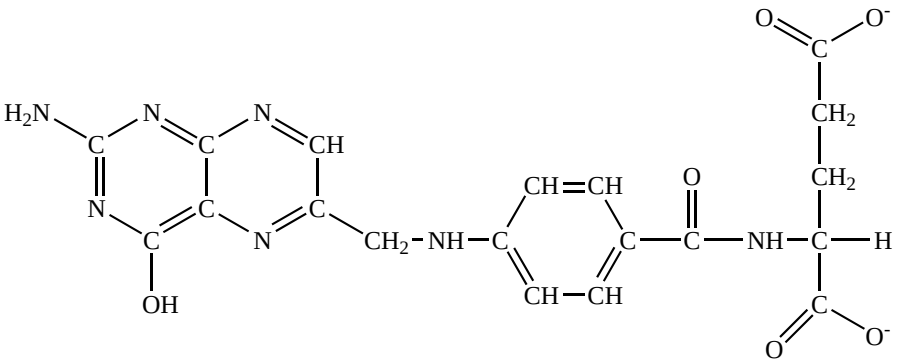
Folic acid

Bacteria in the intestinal track also produce *folic acid;* however, green leafy vegetables, dried beans, and liver are also sources. Reduction of folic acid yields tetrahydrofolic acid, the active form. Both structures are shown in Figure 10-6. The coenzyme transports a carbon, usually as a methyl or formyl, in the synthesis of heme, nucleic acids, choline, and several other compounds. Although cooking easily destroys the compound, intestinal bacteria normally produce sufficient quantities.

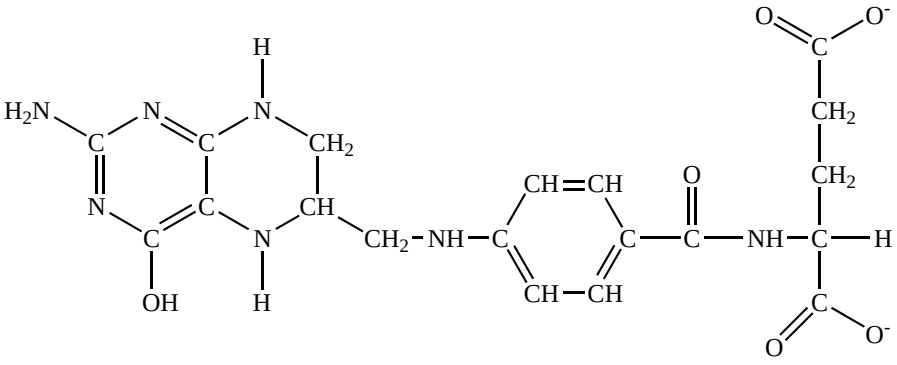
Folic acid is critical to the prevention of malformations of the brain (*anencephaly*) and spine (*spina bifida*). A deficiency of folic acid affects the synthesis of purines — symptoms include gastrointestinal disturbances and anemia. Pregnant women are normally advised to take a vitamin high in folic acid to help in the normal development of the fetus, especially the spine and brain. Sulfa drugs interfere with the formation of folic acid by some pathogens via a form of competitive inhibition.

Pantothenic acid

Pantothenic acid's name derives from a Greek word meaning "from everywhere." As you might expect, then, it has numerous sources, including whole grains, eggs, and meat. Deficiency is virtually unknown. The vitamin is not destroyed by moderate cooking temperatures, but it is not stable at high cooking temperatures. Its structure appears in Figure 10-7.



Folic acid



Tetrahydrofolic acid

Figure 10-6: Structures of folic acid and tetrahydrofolic acid.

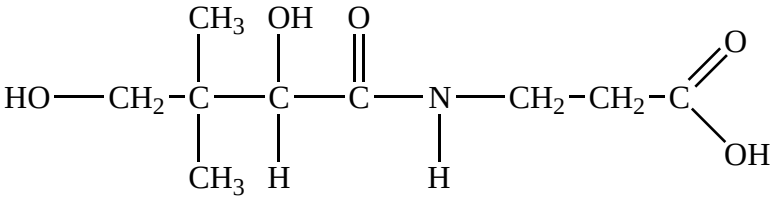


Figure 10-7: Structure of pantothenic acid.

Pantothenic acid is necessary in the biosynthesis of coenzyme A. Coenzyme A is an exceedingly important substance in many biological processes because this coenzyme transfers acyl groups.

The wonders of vitamin B_{12}

Vitamin B_{12} is the only known natural organometallic compound. It does not occur in higher plants, and apparently only bacteria are capable of synthesizing it — bacteria who live in their hosts in a symbiotic relationship. Unfortunately, higher animals including human beings do not have these types of bacteria. Thus, it is necessary to obtain vitamin B_{12} from food. The name *cyanocobalamine* refers to the presence of cyanide. The cyanide is an artifact of the isolation of the compound and is not naturally present. Vitamin B_{12} is necessary to the formation of two coenzymes: methylcobalamin and 5'-deoxyadenosylcobalamin. The structure of methylcobalamin is shown in Figure 10-8.

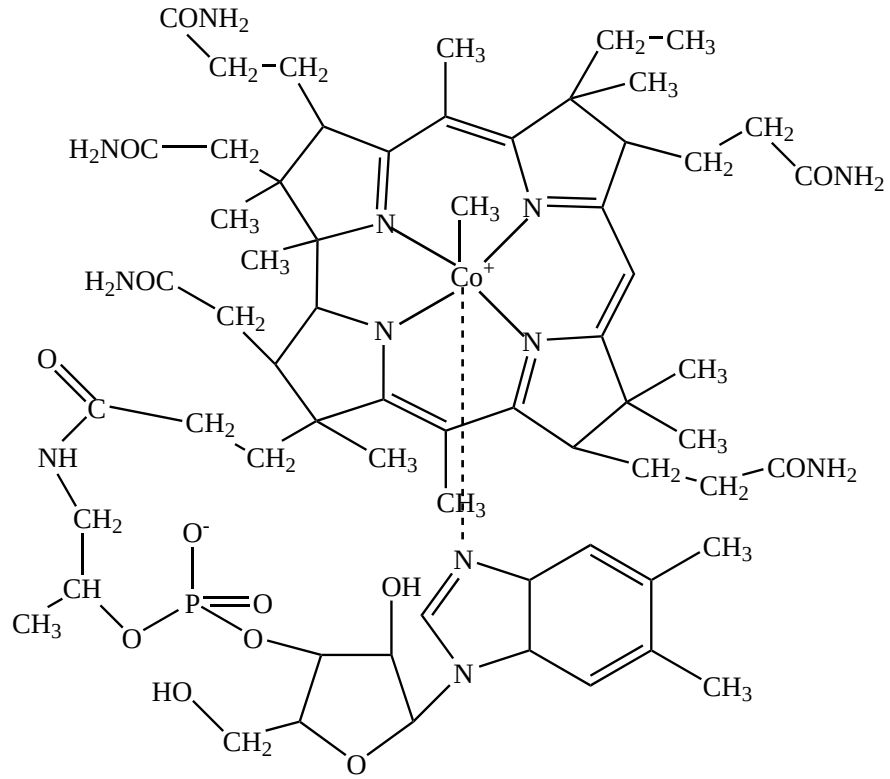


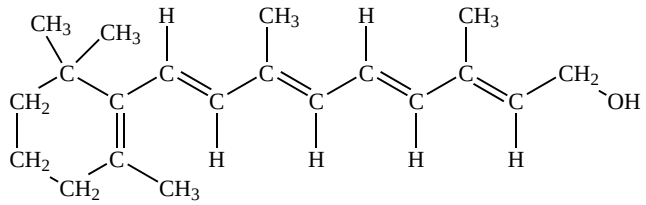
Figure 10-8: Structure of methylcobalamin.

Both coenzymes assist in reactions involving rearrangements. Methylcobalamin is useful in methyl transfer reactions. The coenzyme 5'-deoxyadenosylcobalamin works in some rearrangement reactions where a hydrogen atom and a group attached to an adjacent carbon exchange positions.

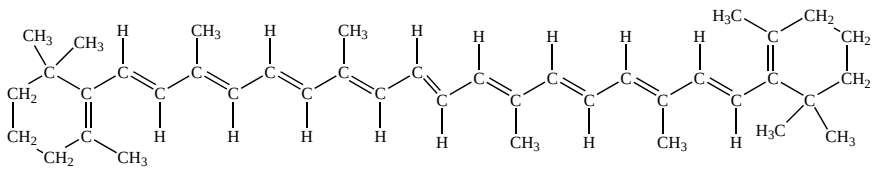
Pernicious anemia usually results from poor absorption of vitamin B_{12}. Normal stomach cells produce a glycoprotein that aids in the absorption of the vitamin in the intestine. It is the lack of this intrinsic factor that leads to the vitamin deficiency and not the lack of the vitamin in the diet. Elderly people may have difficulty in generating sufficient quantities of the intrinsic factor, and strict vegetarians also may develop symptoms. The symptoms of pernicious anemia include lesions on the spinal cord leading to a loss of muscular coordination and gastrointestinal problems. The blood contains large, fragile, and immature red blood cells. Dietary sources include meat, eggs, milk and cereals. This vitamin is stable to cooking.

Vitamin A

Vitamin A is not a single compound — a number of compounds are biologically active, that is they undergo biological reactions within the organism. The parent compound is 11-trans-retinol, found in milk and eggs. Vitamin A is exclusive to animals, and the plant pigment β-carotene can serve as a precursor (Figure 10-9). As a precursor, it is a *provitamin.* Cleavage of β-carotene yields two vitamin A active species. Any β-carotene that doesn't become vitamin A is used as an antioxidant.



11-trans-retinol



β-carotene

Figure 10-9: Structures of 11-trans-retinol and β-carotene. Note that carbon 11 is the fifth from the right in the main chain.

Vitamin A is especially important to vision. Part of the vision process involves the absorption of light. This absorption causes the geometry on the double bond between carbon atoms 11 and 12 to change from cis to trans. The isomerization triggers a series of events, giving rise to a nerve impulse. An enzyme reverses the isomerization so the molecule may be reused. In addition to being directly involved in vision, vitamin A also promotes the development of the epithelial cells producing the mucous membranes, which protect the eyes and many other organs from infections and irritants. Vitamin A also helps in the changes in the bone structures that occur as an infant matures.

A deficiency in vitamin A begins with night blindness, followed by other eye problems, which could lead to blindness. An extreme deficiency may lead to *xerophthalmia,* inflammation of the eyelids and eyes, which can cause infections and blindness. Young animals require vitamin A for growth, and adults are capable of storing several months' supply of it, primarily in the liver. The livers of some animals, such as polar bears and seals, may have such a high vitamin A concentration that they are toxic to humans. Excessive dosages of vitamin A may lead to acute toxicity, and as a fat-soluble vitamin, it is not easily eliminated. Symptoms include nausea, vomiting, blurred vision, and headaches. Large doses have been linked to birth defects and spontaneous abortions. The provitamin, β-carotene, is not toxic.

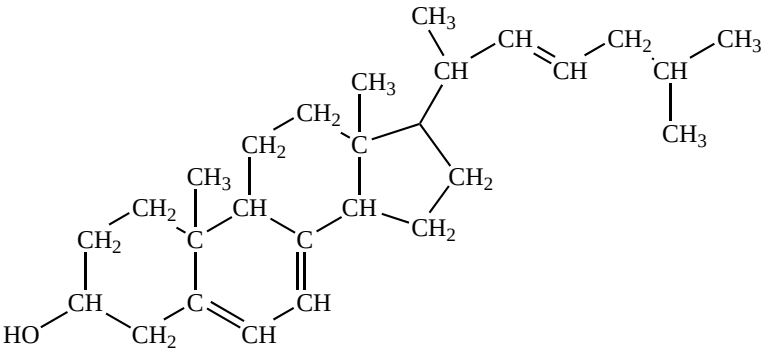
Vitamin D

Vitamin D is sometimes called the *sunshine vitamin*. It can be produced in the body through the action of sunlight, which is ultraviolet radiation. Individuals walking around outside nude or semi-nude normally have very little trouble with vitamin D deficiency. The rest of us, however, depend on vitamin D-fortified foods, especially milk.

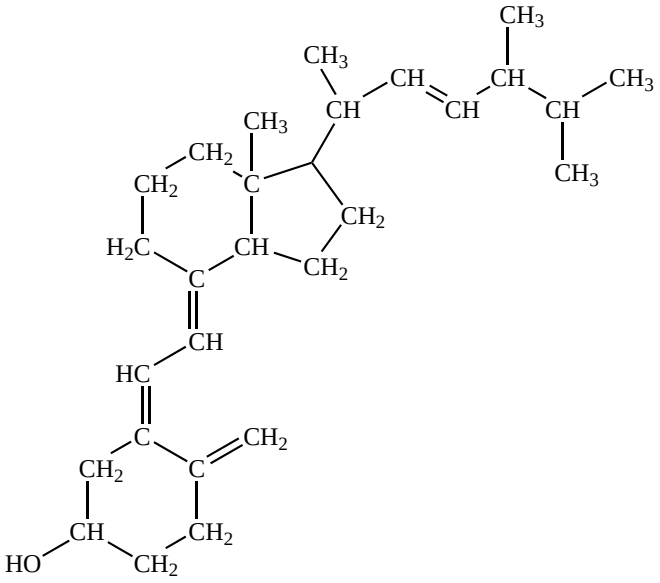
Several compounds exhibit vitamin D activity. Only two of them — actually provitamins — occur commonly in food: ergosterol and 7-dehydrocholesterol. Irradiation with ultraviolet light converts ergosterol into vitamin D_2, *ergocalciferol.* Ultraviolet irradiation, particularly in the skin of animals, converts 7-dehydrocholesterol into vitamin D_3, *cholecalciferol.* (A little confusingly, vitamin D_1 is a mixture of vitamin D_2 and vitamin D_3.) The structures of ergosterol, vitamin D_2, 7-dehydrocholesterol, and vitamin D_3 appear in Figure 10-10.

The body's ability to absorb calcium and phosphorus is tied to vitamin D. Teeth and bone have large amounts of these two elements and are the first parts of the body affected by a vitamin D deficiency. Osteomalacia, a condition in which a softening of the bones may lead to deformities, may also result. (In infants and children, osteomalacia is called rickets.) A vitamin D

deficiency is more serious in children than in adults because growth requires larger quantities of calcium and phosphorus. Persons with some portion of their skin routinely exposed to sunlight seldom develop a deficiency.

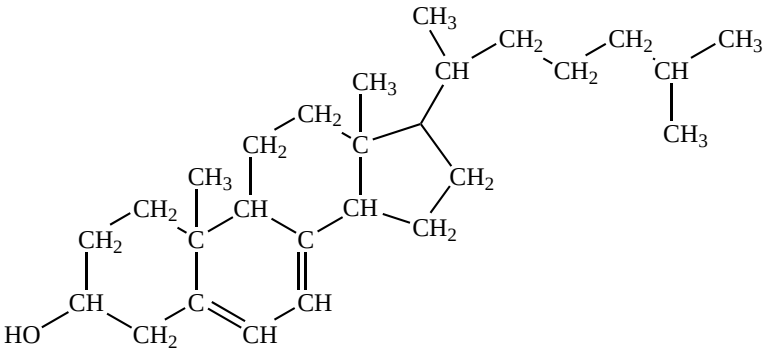


Ergosterol

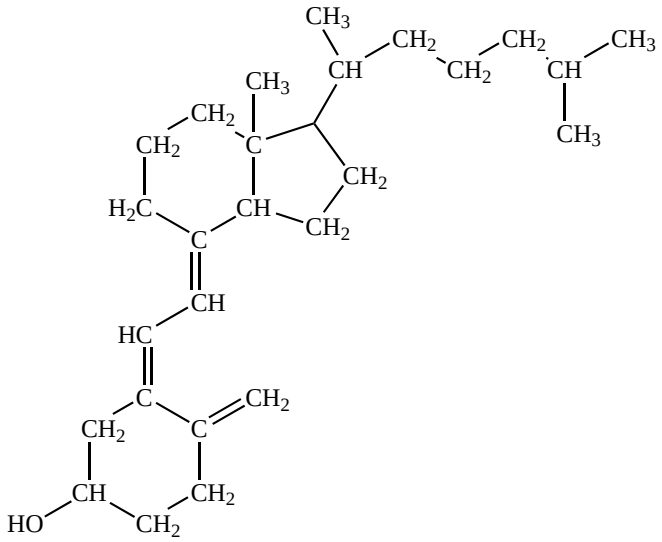


Vitamin D_2

Figure 10-10: Structures of ergosterol, vitamin D_2, 7-dehydrocholesterol, and vitamin D_3.



7-Dehydrocholesterol



Vitamin D_3

Figure 10-10: (continued)

Excess vitamin D is toxic. It is not easy to eliminate this fat-soluble vitamin. Symptoms of excessive amounts of vitamin D include nausea, diarrhea, kidney stones and other deposits, and sometimes even death.

Vitamin E

The *tocopherols* are a group of compounds that exhibit vitamin E activity. The most effective is α-tocopherol (see Figure 10-11). Vitamin E comes from a number of sources, vegetable oils, nuts, whole grains, leafy vegetables, to name a few. Deficiencies are rare except in individuals on a no-fat diet or who, for medical reasons, cannot efficiency absorb fat. Cystic fibrosis may interfere with fat absorption.

Vitamin E serves as an effective anti-oxidant. *Anti-oxidants* are necessary to minimize the damage caused by oxidants present in the body — many problems associated with aging are apparently due to oxidants. Vitamin E may also help prevent cholesterol deposits in the arteries. There are no well-documented problems with the use of large doses; however, some recent studies, although disputed, warn against taking mega doses of vitamin E.

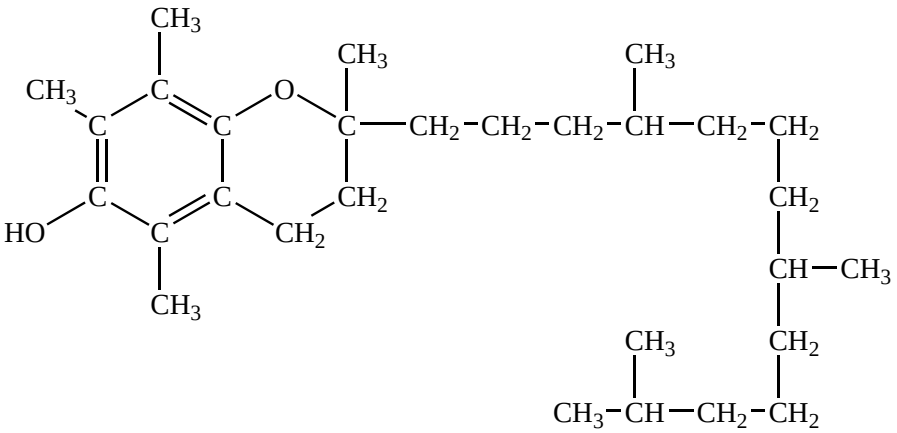


Figure 10-11: Structure of α-tocopherol (vitamin E).

Vitamin K

Vitamin K_1 (Figure 10-12) is one of many compounds that exhibit vitamin K activity and differ in the side-chains attached to the ring system. One chain is usually a methyl; the other typically has at least 20 carbon atoms.

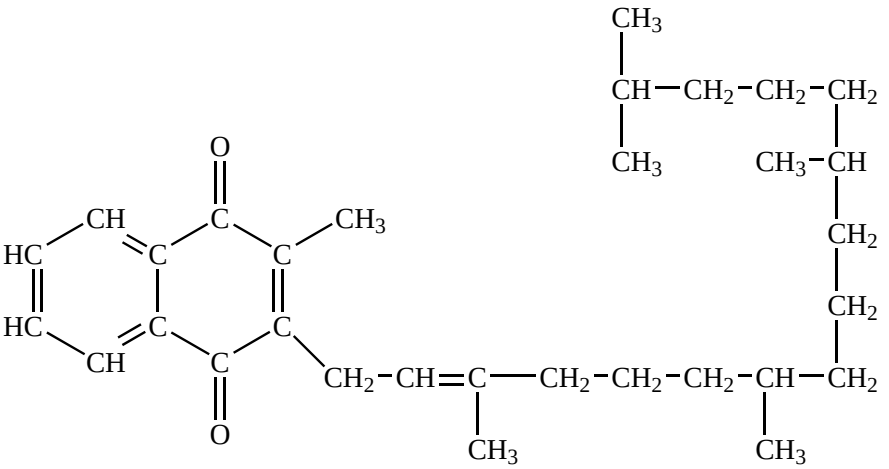


Figure 10-12: Structure of vitamin K_1.

Vitamin K is necessary to produce the proenzyme *prothrombin,* which helps blood clot. A vitamin K deficiency is uncommon because intestinal bacteria normally produce sufficient quantities, although several foods are also good sources, including green leafy vegetables, cauliflower, broccoli, organ meats (love that liver!), milk, soybeans, avocados, and bananas. Two tablespoons of parsley contains almost twice your recommended daily amount of vitamin K. Prolonged use of antibiotics can decrease the number of these vitamin K-producing bacteria and lead to a reduction in vitamin K in the body. One symptom of a deficiency is an increase in the time necessary to form a blood clot, and such individuals are prone to develop serious bruises from even minor injuries. Infants with a deficiency have been known to die from brain hemorrhage. Increasing the vitamin K intake of the mothers decreases the likelihood of this occurrence.

Vitamin C

Vitamin C is another name for *ascorbic acid* (Figure 10-13). Dehydroascorbic acid also has vitamin C activity. Vitamin C is water-soluble — thus the body can readily eliminate excess, and large doses are not toxic.

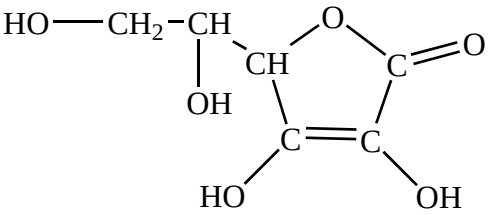


Figure 10-13: Structure of vitamin C.

A deficiency in vitamin C leads to the disease *scurvy,* symptoms of which include a weakening of the collagen — an important protein in connective tissues such as ligaments and tendons. Many foods contain vitamin C, especially plants and citrus fruits, so it is easy to prevent scurvy. For years, British ships carried limes as a source of vitamin C (leading, incidentally, to the slang term *limey* to refer to a British sailor). Many mammals (other than humans) synthesize vitamin C from glucose. Cooking, especially prolonged cooking, destroys vitamin C. Vitamin C is an antioxidant. Like vitamin E, it helps prevent damage produced by oxidants. It also helps in the absorption of iron, and keeps the iron in the +2 state. Vitamin C helps convert some of the proline in collagen C to hydroxyproline, which stabilizes the collagen.

Chapter 11

Be Quiet: Hormones

In This Chapter

- Examining the structures of hormones
- Finding out about some important hormones
- Discovering how hormones function

We *know* with this chapter title you are expecting several jokes, but we'll try to restrain ourselves and play it straight.

Hormones are materials produced in one area of the body and used in a different area. They are molecular messengers that are created in certain glands in the body and then travel through the bloodstream to the target organ. Other substances, called *paracrine factors* or *growth and differentiation factors* (GDFs), also convey biochemical information within a particular organ (a bit like passing a note in class). This conveyance is accomplished by simple diffusion over a small distance. Some biochemical substances may be both a hormone and a paracrine factor.

The endocrine glands produce most — but not all — hormones. Endocrine glands include the hypothalamus, pituitary, pancreas, adrenal, liver, testes, and ovaries. Now surely that got your attention! Some glands produce a single hormone, whereas others produce more than one. The simplified viewpoint (and we are all about keeping it simple) is that the pituitary gland acts as the central control for the endocrine system. Hormones from the pituitary gland do cause other glands to produce hormones that affect other systems — however, there are glands that have the same effect on the pituitary gland.

Structures of Some Key Hormones

There are three groups of hormones:

- Proteins, such as insulin
- Steroids, materials derived from cholesterol
- Amines, such as epinephrine

These materials allow one part of the body to influence what occurs elsewhere. These molecules are so efficient that only very low concentrations, typically 10^{-7} to 10^{-10} M, are necessary. That's a really small amount! The low concentrations make identification and isolation of these substances difficult.

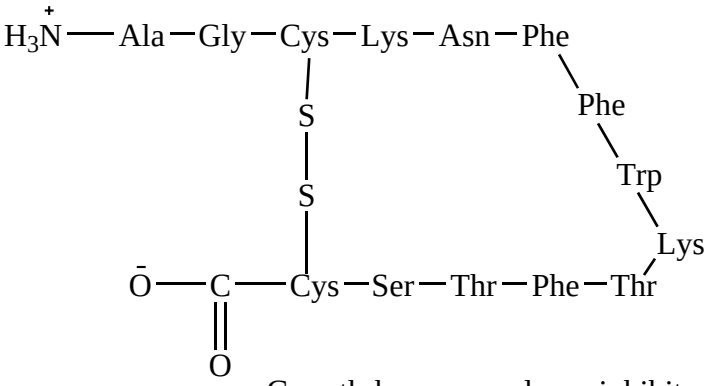
Proteins

The *protein,* or *polypeptide,* hormones, typically produced by the pituitary and hypothalamus glands, vary greatly in size — from simple tripeptides to larger molecules with more than 200 amino acid residues. Protein hormones are a diverse collection of molecules, including insulin (the structure of which you can see in Chapter 5).

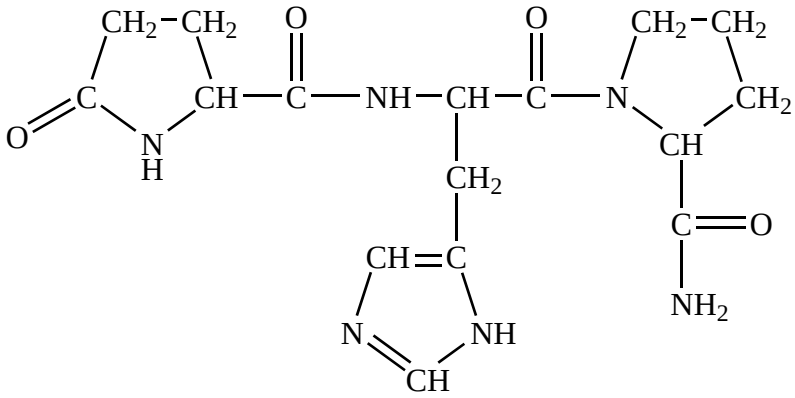
Others include the *thyrotropin-releasing factor,* which induces the release or production of a biochemical (thyrotropin, in this case). The thyrotropin-releasing factor hormone is a tripeptide containing glutamine (modified), histidine, and proline (modified). Another one is the *growth-hormone-release-inhibitory factor,* which inhibits the release or production of a chemical species. Together (Figure 11-1), these types of hormones provide a mechanism to start and stop an action. The idea is to maintain a tight biochemical control of biochemical processes, such as growth.

Steroids

You have no doubt read about steroid use among athletes, where it is used to increase muscle mass — to "pump-up," in other words. *Steroid* hormones, produced by the body's ovaries, testes, and adrenal glands, are cholesterol derivatives of about the same size as the parent molecule. They include the *estrogens* (female sex hormones), the *androgens* (male sex hormones), and the *adrenal cortical* hormones, such as aldosterone and cortisol. The estrogens and androgens are responsible for the development of the secondary sex characteristics of both females and males, respectively. These characteristics include enlargement of the breasts of females and development of facial hair in males.



Growth-hormone-release-inhibitory factor



Thyrotropin-releasing factor

Figure 11-1: Structures of the growth-hormone-release-inhibitory factor and thyrotropin-releasing factor.

The adrenal cortical hormones (Figure 11-2), which include the glucocorticoids and the mineralocorticoids, have a variety of functions. The glucocorticoids, such as cortisol, are important to several metabolic pathways. The mineralocorticoids, such as aldosterone, are important to the transport of inorganic species, such as sodium or potassium ions.

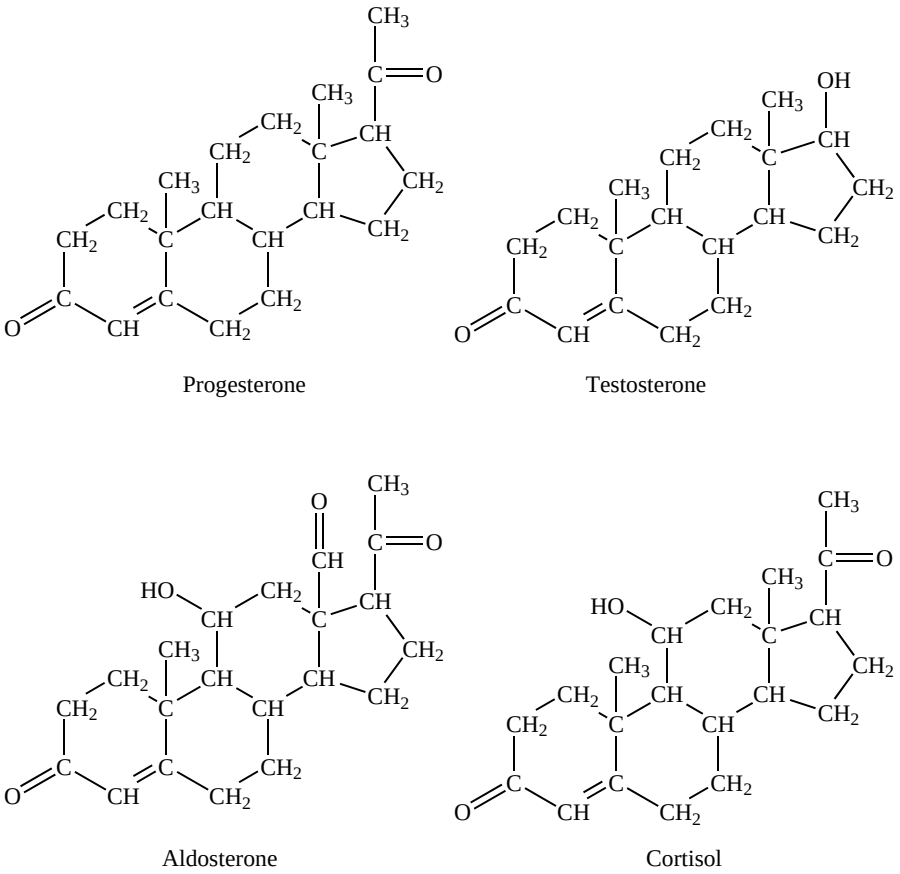
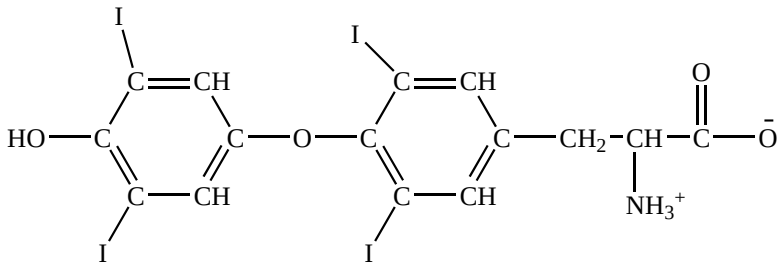


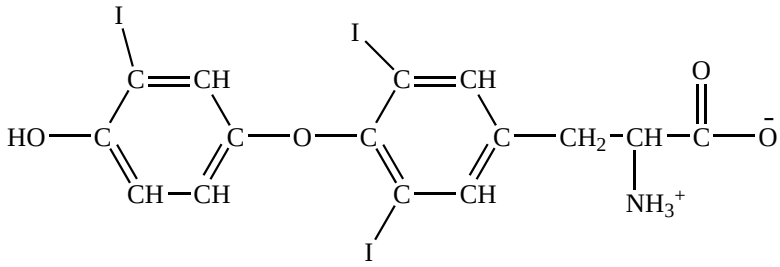
Figure 11-2: Structures of progesterone (an estrogen), testosterone (an androgen), and aldosterone and cortisol, adrenal cortical hormones.

Amines

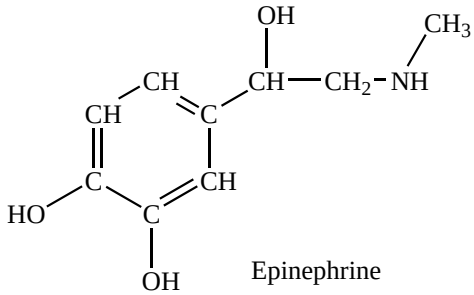
The *amine* hormones, typically produced by the thyroid and adrenal glands, are small molecules, many of which are derivatives of tyrosine. These hormones include *thyroxine* and *triiodthyronine,* produced by the thyroid gland, and *epinephrine* and *norepinephrine* produced by the adrenal gland. Figure 11-3 illustrates the structures of these hormones. Thyroxine and triiodothyronine are important metabolic-rate regulators. In fact, thyroxine is one of the most important substances in the body. It influences carbohydrate metabolism as well as protein synthesis and is involved in cardiovascular, brain, and renal function. Epinephrine and norepinephrine are important to control heart rate, blood flow, and metabolic rate.



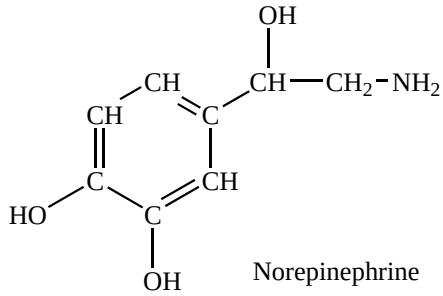
Thyroxine



Triiodothyroxine



Epinephrine



Norepinephrine

Figure 11-3: Structures of thyroxine, triiodothyronine, epinephrine, and norepinephrine.

Before and After: Prohormones

The synthesis of some hormones, like some enzymes, does not begin by producing the molecule in its active form. Instead, a *prohormone* forms, which remains unreactive and dormant until activated — sort of like us in the morning until we get our first cups of coffee. This process allows the body to build a store of a hormone for quick activation. Activating the prohormone requires less time than would the total synthesis of the molecule.

Proinsulin

Proinsulin is an example of a prohormone. *Insulin* is the hormone responsible for controlling blood sugar levels. Too much insulin results in a low blood sugar level (hypoglycemia), whereas too little insulin leads to elevated blood sugar levels (hyperglycemia). Your body needs to have a supply of insulin readily available for when you eat a piece of candy, such as a large chocolate-hazelnut truffle. If all of this insulin were already in your bloodstream, upon eating the candy you would become hypoglycemic. If the insulin were not present at all, you might become hyperglycemic until your body was able to synthesize sufficient insulin from the individual amino acids. Both hypoglycemia and hyperglycemia can lead to serious medical problems. The presence of a quantity of inactive insulin, ready to jump into action at a moment's notice, is the solution.

Bovine insulin (insulin produced from cows) contains two polypeptide chains, A and B, linked by disulfide linkages, with a total of 51 amino acid residues. Bovine proinsulin has 30 more amino acid residues than insulin does. Proinsulin is a single polypeptide chain with the insulin disulfide linkages. By removing a polypeptide sequence from the central region of this chain (residues 31–60), insulin forms. The excised portion originally connected one end of the A chain of insulin to the end of the B chain. The conversion of proinsulin to insulin requires the cleavage of two peptide bonds.

Angiotensinogen

Angiotensinogen is the prohormone of angiotensin II, a hormone that signals the adrenal cortex to release aldosterone. (In addition, angiotensin II is the most potent known vasoconstrictor.) The conversion of the prohormone to the hormone requires two steps. The first step uses the enzyme rennin. This peptidase, produced in the kidney, specifically cleaves a peptide bond between two leucine residues, the result of which is the decapeptide angiotensin I.

The second step utilizes the peptidase known as the angiotensin-converting enzyme. This enzyme, which occurs primarily in the lungs, cleaves the C-terminal dipeptide from angiotensin I to yield angiotensin II. These biochemical reactions can occur very rapidly, ensuring that the hormone can be quickly activated when needed by the body.

Fight or Flight: Hormone Function

The endocrine system, which generates the hormones, consists of a number of apparently unrelated organs: the liver, the ovaries or testes, the thyroid, the pancreas, and a number of other glands — components that are part of a complex, integrated network. A malfunction of one affects others.

Opening the letter: Hormonal action

Several mechanisms lead to the regulation of hormones. A *control loop* is the simplest. In many cases, one hormone stimulates the production of others so that many actions may occur before some type of control occurs.

Simple control loops

We are all familiar with *control loops.* You study for a test, but get a so-so grade. So you study harder for the next exam. Your grade provides *feedback,* causing your study habits to (hopefully) change. In the body, a control loop process begins with an external stimulus signaling a gland to generate a hormone. This hormone then influences its target site. Action by the target leads to a change, which signals the gland to stop. The action of the hormone causing the stop signal provides a negative feedback. An example of this type of loop is the production of insulin by the pancreas. The presence of high glucose levels in the bloodstream signals the pancreas to release insulin. The released insulin lowers the glucose level in the bloodstream. The reduced glucose level signals the pancreas to stop releasing insulin. The low glucose level is the negative feedback. This is a simplification; other factors may trigger the release of insulin. In addition, high glucose levels can trigger other biochemical functions, such as the synthesis of glycogen in the liver.

Hypothalamus-pituitary control

The hypothalamus-pituitary system is a very complex example of the other extreme of hormone control. The hypothalamus and the pituitary glands are in such close proximity that they behave almost as a single unit.

Initially, the central nervous system signals the hypothalamus to release a hormone called a *hormone-releasing factor,* which signals the pituitary. The pituitary, thus signaled, releases another hormone into the bloodstream. This hormone may target a specific organ or signal another part of the endocrine system to secrete yet another hormone. The presence of this final hormone serves as a negative feedback signal to the hypothalamus to stop secreting the hormone-releasing factor to the pituitary. Again, this is a simplistic view of a complicated system. An analogy might be your parents seeing your so-so exam grade. They freak out and force you to study harder. You are being influenced by an external force, in this case, your parents.

Figure 11-4 gives a more detailed representation of this system. Although the pituitary gland is known as the "master gland," this figure indicates that it is, in fact, the hypothalamus that deserves this honor.

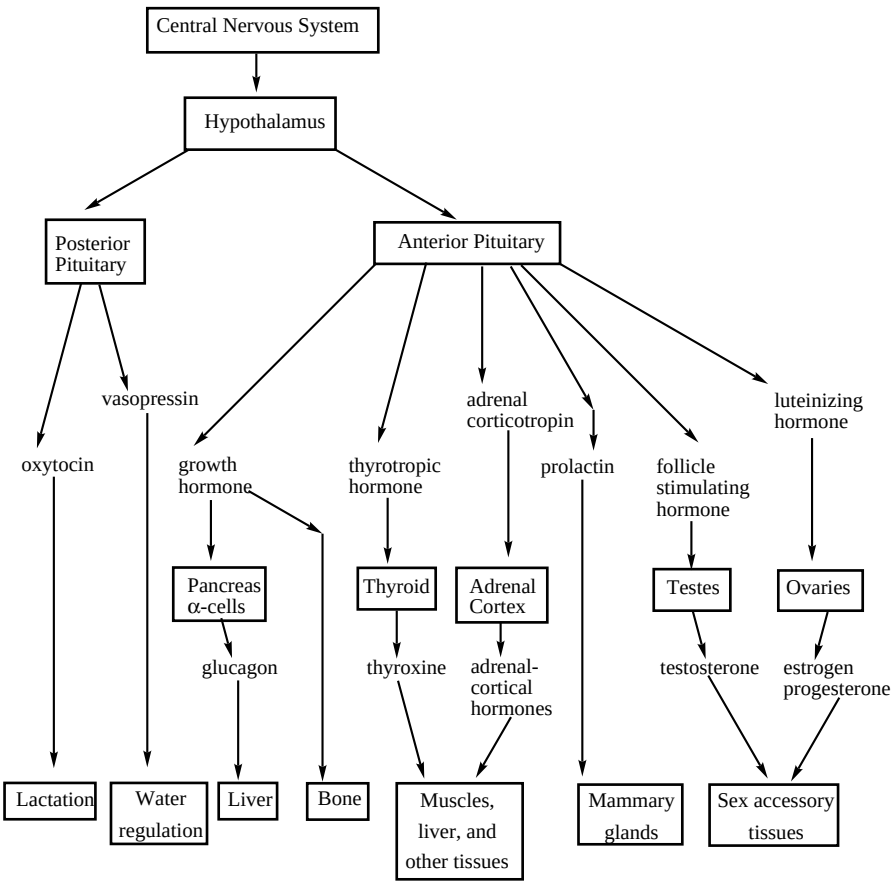


Figure 11-4: Schematic of hormone control in the body.

Models of hormonal action

Two models have been proposed to account for the molecular action of hormones. The first is the *two-messenger hypothesis,* which applies primarily to polypeptide and amine hormones. The other, *steroid hormonal action,* applies primarily to steroids. We use a simplistic approach (the KISS rule: Keep It Simple, Silly) for each model to emphasize their basic concepts. The actual processes involve many more changes.

The two-messenger model: Like the mail

Studies into the hormonal action of epinephrine (adrenaline) led to the development of this model. Later work indicated that the model applies to other hormonal systems as well. In the two-messenger model, a hormone binds to a receptor site on the exterior of a cell. This binding induces the release of another agent within the cell. The hormone is the first messenger, and the other agent in the second messenger.

For example, the adrenal medulla releases epinephrine, the "fight or flight" hormone, in vertebrate animals. This release initiates a number of responses, including glycogenolysis, the breakdown of glycogen. Glycogenolysis releases glucose for use in rapid energy production. As with other hormones, the concentration of hormone required is very low. For epinephrine, it is about 10^{-9} M. The released epinephrine acts as the first messenger (the extracellular one). Molecules of epinephrine bind to specific receptor sites on the surface of the target cells — primarily the skeletal muscles and, to a lesser extent, liver. The binding of epinephrine to the outside of liver cells induces the enzyme adenylate cyclase, bound to the interior of the cell membrane, to synthesize cyclic AMP (see Figure 11-5). Cyclic AMP, or cAMP, is the second messenger (the intracellular one). The second messenger initiates a series of events terminating in the release of glucose (glycogenolysis).

Initially, the cAMP binds to the regulatory subunit of protein kinase, and this activates the membrane-bound enzyme. The released protein kinase then activates phosphorylase kinase. This process requires calcium ion and ATP. (Muscular action releases calcium ion, which aids the process.) Phosphorylase kinase, with aid of ATP and magnesium ion, converts inactive phosphorylase b to active phosphorylase a. The increased presence of this enzyme accelerates the breakdown of glycogen with the release of D-glucose-1-phosphate. Phosphoglucomutase then converts D-glucose-1-phosphate to D-glucose-6-phosphate. Finally, D-glucose-6-phosphatase catalyzes the loss of the phosphate to release glucose, which may be used in the cell or, more importantly, may enter the bloodstream. Whew!

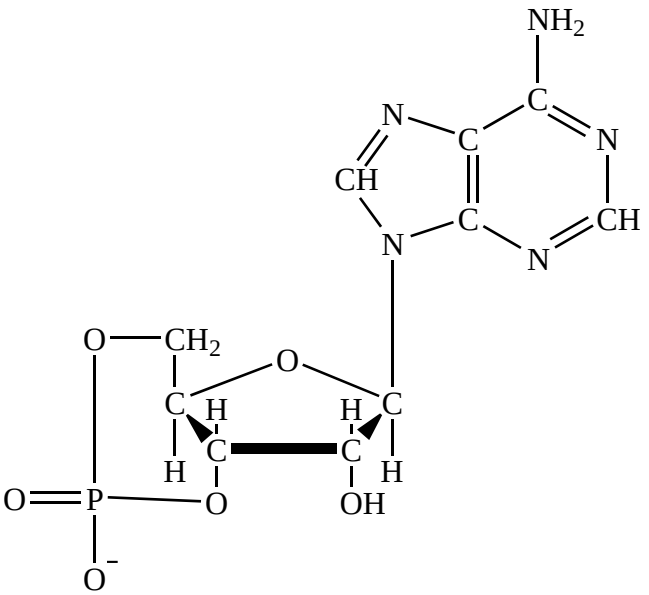


Figure 11-5: Structure of cyclic AMP.

The enzyme protein kinase also catalyzes the conversion of glycogen synthase (active) to phosphor-glycogen synthase (inactive). Thus, while the level of protein kinase is high, the production of new glycogen ceases. The inhibition of glycogen synthesis also means that more glucose will be available for rapid actions, such as running away from an angry pit bull.

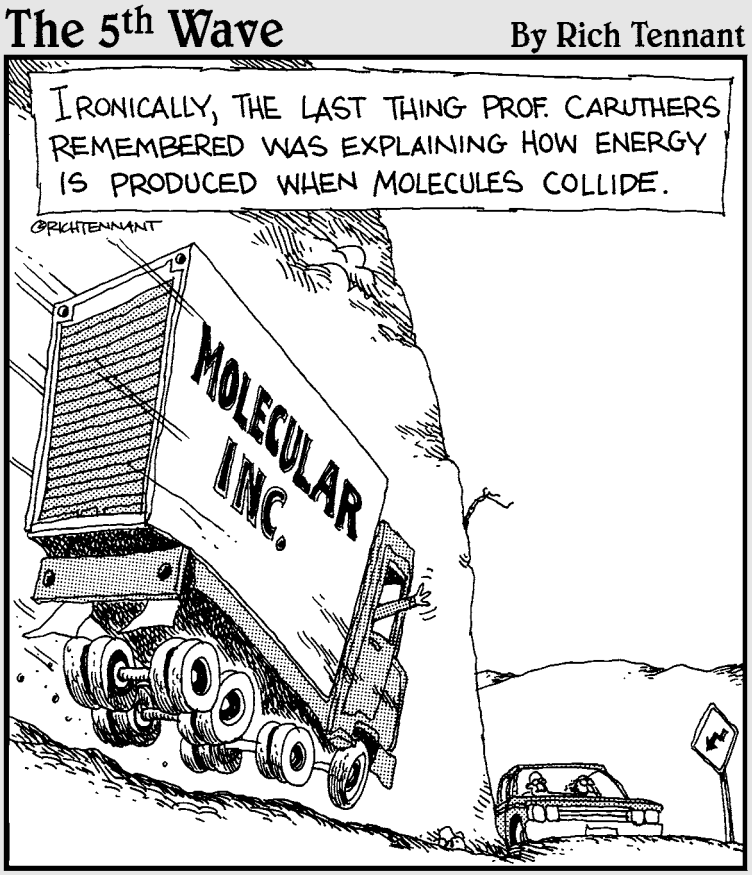
Steroid hormonal action

Unlike hormones in the two-messenger system, steroid hormones cross the membrane and enter the cell. This mechanism applies to other hormones as well, such as thyroid hormones, in addition to the steroid hormones.

The first system described by this model was the action of estradiol on uterine tissue in mammals. The estradiol, an estrogen, enters the cell where it binds to an estrogen-receptor protein. The binding does not involve covalent bonding, but induces instead a conformational change in the protein. The change in the shape of the protein allows it to pass through the "door" into the cell nucleus. The hormone-protein complex then enters the cell nucleus where it binds to a specific site on a chromosome. This binding to the chromosome stimulates transcription to produce mRNA, which, in turn, exits the nucleus and synthesizes protein molecules through translation.

Three basic factors differentiate the steroid system from the two-messenger system. First, in the steroid system the hormone enters the cell. Second, there is a specific receptor molecule within the cytosol, the fluid inside the cell. Finally, the hormone action is at the chromosome level.

Part IV
Bioenergetics and Pathways



In this part . . .

For anyone to do anything requires energy, and this is where we focus on the way life obtains and uses it. Here we take a gander at energy needs and follow the trail of where that energy goes and why. The main character in this part is your good buddy ATP, and running through this episode is where you'll find the citric acid cycle. Finally, we tackle nitrogen chemistry.

Chapter 12

Life and Energy

In This Chapter

- Learning about ATP and energy
- Visiting the nucleoside triphosphate family
- Considering AMP, ADP, and ATP
- Going without food

The chapters in this part examine metabolism — all the processes involved in maintaining a cell. Metabolism has two components: catabolism and anabolism. *Catabolism* deals with the breaking down of molecules, whereas *anabolism* deals with the building up of cells. Both processes take place in the mitochondria. All metabolic processes involve energy: They either absorb energy (*endergonic*) or produce it (*exergonic*).

The key energy molecule is *adenosine triphosphate,* abbreviated *ATP,* which forms as a product of the common catabolic pathway.

ATP: The Energy Pony Express

Determining the basic reaction processes involved in the production and use of energy is called *bioenergetics*. This study has developed bioenergetic principles that allow us to examine energy at the microscopic level.

Fortunately, ATP is recycled within the body. The typical daily requirement for an adult is over 140 pounds of ATP per day. However, the amount of ATP present in your body at any one time is only about one-tenth of a pound. That means each ATP molecule in your body is recycled about 1,400 times each day. Now that is effective recycling — and you don't even have to put anything into a blue container.

ATP and free energy

The *free energy content* (G) is the intrinsic energy present in a molecule. In a reaction, the change in this energy is written as ΔG. The change in energy is equal to the energy of the products minus the energy of the reactants. The value of ΔG is the key: If a reaction produces energy, ΔG represents the maximum possible amount of energy that the reaction may produce. If a reaction requires energy, ΔG represents the minimum possible amount of energy that a reaction will require. Reactions producing energy have a negative value of ΔG and are *spontaneous.* Reactions requiring energy have a positive value of ΔG and are *nonspontaneous.*

Spontaneity bears no relation to speed. Spontaneous reactions may be very rapid or very slow.

The conditions under which a reaction occurs may alter the value of ΔG. (The "ideal" or standard value of ΔG is ΔG°.) The formula for modifying the free energy for the equilibrium reaction A ⇌ B is:

$$\Delta G = \Delta G^\circ - RT \ln [B] / [A] = \Delta G^\circ - RT \ln K$$

According to this relationship, the free energy change, ΔG, comes from a modification of the standard free energy value. R is the universal gas constant ($8.314\ J \times mol^{-1}K^{-1}$ or $1.987\ cal \times mol^{-1}K^{-1}$). T is the absolute temperature. K is the equilibrium constant found by dividing the concentration of the product, [B], by the concentration of the reactant, [A].

In many bioenergetic studies, *calories* are the unit instead of joules (J). The relationship is 1 calorie = 4.184 J (exactly) or 1 kilocalorie = 4.184 kJ.

In research, it is often better to use ΔG°'. This modification of ΔG stems from the use of the biologically more realistic value of pH = 7 ($[H^+] = 10^{-7}$ M) instead of the standard pH = 0 ($[H^+] = 1$ M). Some relationships between K and ΔG°' are shown in Table 12-1.

Table 12-1 Relationships Between ΔG°' and K

$\Delta G^{\circ\prime}\ kJ \times mol^{-1}$	K
–17.1	1,000
–11.4	100
–5.7	10
0	1

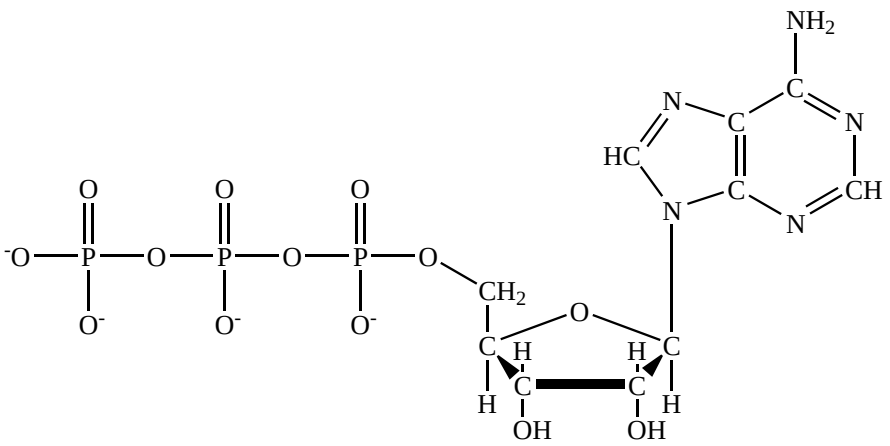
$\Delta G^{\circ\prime}$ kJ × mol^{-1}	K
5.7	0.1
11.4	0.01
17.1	0.001

Table 12-1 shows that the larger K is, the more *exergonic* (spontaneous) the reaction. For example, if K = 1000, the concentration of the product, [B], is 1,000 times that of the reactant, [A], and 17 kJ per mole will be released. It is important to remember that, in biological systems, variations in [A] and [B] must be taken into account in addition to $\Delta G^{\circ\prime}$. For example, increasing the reactant concentration promotes the reaction, whereas increasing the product concentration inhibits the reaction.

ATP as an energy transporter

Cells utilize exergonic processes to provide the energy necessary for life processes, and the key supplier of this energy is ATP (Figure 12-1). ATP supplies the energy required to force endergonic reactions to take place, to provide mechanical energy (muscle movement), light energy (in fireflies), and heat energy (to maintain body temperature).

Hydrolysis of the terminal phosphate of ATP yields ADP and inorganic phosphate, indicated as P_i. The structure of ADP is shown in Figure 12-2. This hydrolysis releases 30.5 kJ × mol^{-1}.



Adenosine triphosphate (ATP)

Figure 12-1: Structure of ATP.

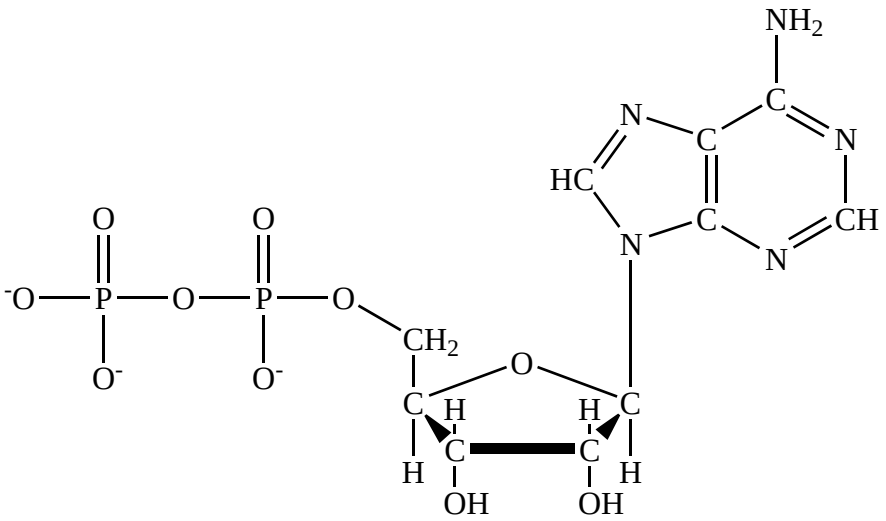


Figure 12-2: Structure of ADP.

Adenosine diphosphate (ADP)

Concentration variations lead to changes, usually minor, in energy.

Hydrolysis of the terminal phosphate of ADP yields AMP and inorganic phosphate, indicated as P_i. The structure of AMP is in Figure 12-3. This hydrolysis also releases 30.5 kJ × mol^{-1}. (This reaction is of less biological importance than the ATP to ADP hydrolysis.)

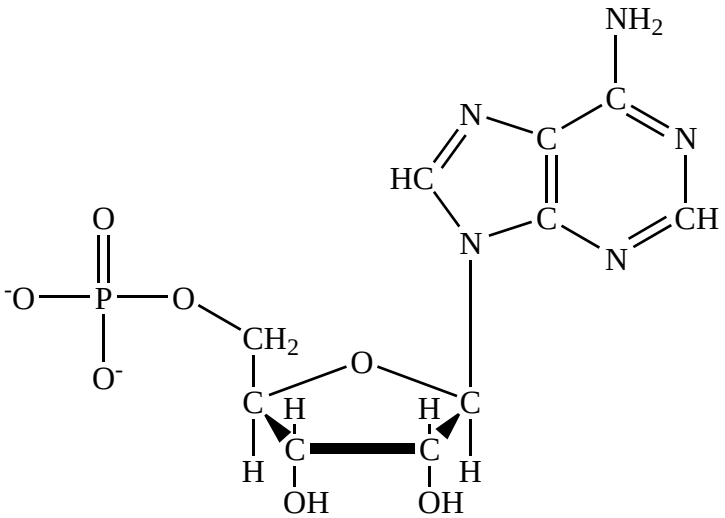
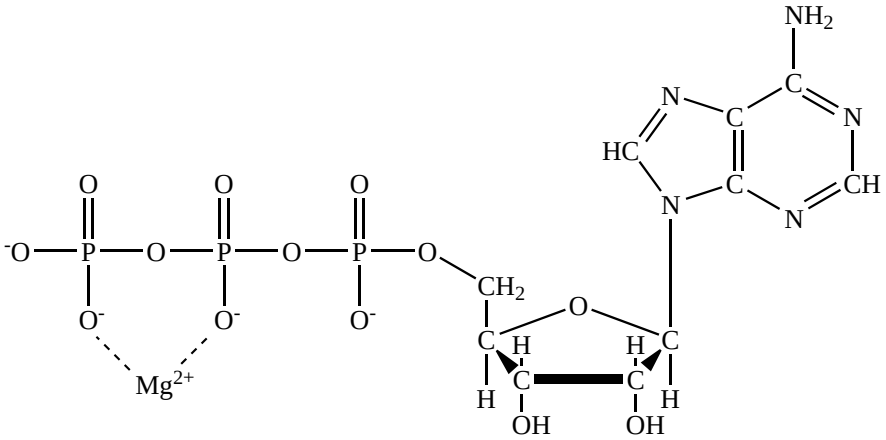


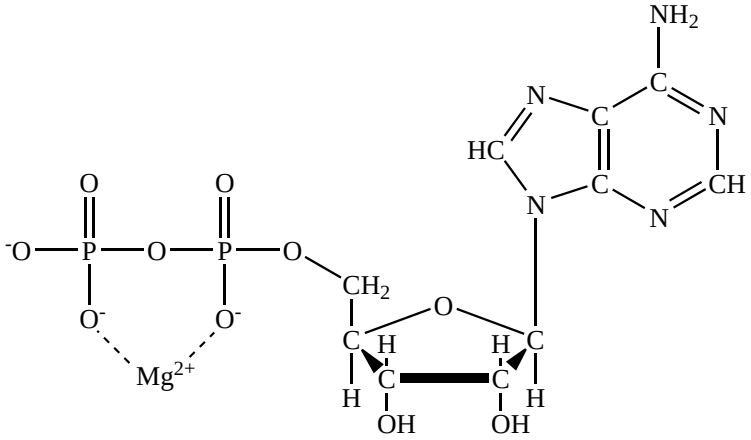
Figure 12-3: Structure of AMP.

Adenosine monophosphate (AMP)

It is also possible to go directly from ATP to AMP, cleaving a pyrophosphate, $P_2O_7^{4-}$, from the phosphate chain. Biochemists use PP_i to indicate pyrophosphate. This furnishes slightly more energy then a simple hydrolysis to release P_i (about 33.5 kJ × mol^{-1}). Under physiological conditions, the phosphate portions of ATP and ADP form a complex with magnesium ions. In certain circumstances, manganese (II) ions, Mn^{2+}, may take the place of Mg^{2+} ions. Figure 12-4 depicts the magnesium complexes with ATP and ADP.



Adenosine triphosphate (ATP)-Mg^{2+}



Adenosine diphosphate (ADP)-Mg^{2+}

Figure 12-4: Magnesium complexes with ATP and ADP.

The removal of the last phosphate involves the loss of the least amount of energy (14.2 kJ•mol^{-1}). This hydrolysis involves the cleavage of an ester bond instead of an anhydride bond. In general, the hydrolysis of an ester bond involves less than half the energy of the hydrolysis of an anhydride bond.

It's Relative: Molecules Related to ATP

A few other biomolecules can provide energy equivalent to that which comes from the hydrolysis of ATP. GTP is an example of such a molecule. There are also a few molecules that supply *more* energy. Table 12-2 compares some of the high-energy molecules to ATP, and Figure 12-5 shows their structures.

Table 12-2 Energy Released ($\Delta G°'$) by Some High-Energy Molecules Related to ATP

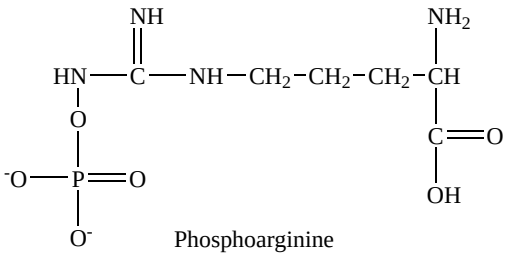
Biomolecule	*Energy released (kJ × mol^{-1})*
ATP	30.5
Phosphoarginine	32.2
Acetyl phosphate	43.3
Phosphocreatine	43.3
1,3-Bisphosphoglycerate	49.6
Phosphoenolpyruvate	62.2

Phosphopyruvate, 1,3-bisphosphogylcerate, and acetyl phosphate are important for the transfer and conservation of chemical energy. Phosphoarginine and phosphocreatine are important molecules for storing metabolic energy. Phosphocreatine is stored in muscles and can be quickly converted to ATP to give energy for muscle contraction. Production of phosphocreatine occurs when ATP concentration is high — high ATP concentration is needed to overcome the energy deficit of 12.8 kJ × mol^{-1}. The reverse, phosphate transfer to form ATP from ADP, occurs at low ATP concentrations. Phosphoarginine behaves similarly in certain invertebrates.

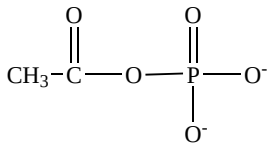
The nucleoside triphosphate family

The predominant energy transfer molecule, as we have been saying, is ATP. But other nucleoside triphosphates (such as CTP, GTP, TTP, and UTP) may also serve this energy transfer function. These five molecules also supply part of the energy necessary for DNA and RNA synthesis. All the nucleoside

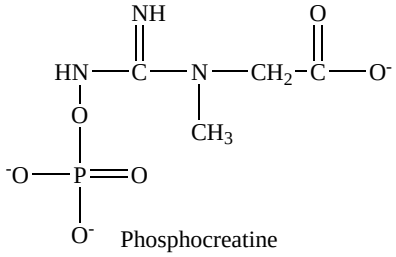
triphosphates have about the same energy yield. (Note that ATP is necessary for the synthesis of the remaining nucleoside triphosphates.)



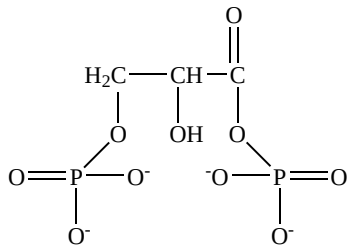
Phosphoarginine



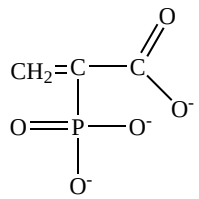
Acetyl phosphate



Phosphocreatine



1,3-Bisphosphoglycerate



Phosphoenolpyruvate

Figure 12-5: Structures of some high-energy molecules.

The biosynthesis of the ribonucleoside triphosphates, in general NTP, begins with the production of the appropriate monophosphate, NMP. The stepwise addition of the next two phosphate groups requires two enzymes of low specificity. These enzymes are nucleoside monophosphate kinase and nucleoside diphosphate kinase. (The term *kinase* refers to a transferase enzyme that transfers a phosphate group of a nucleoside triphosphate.) The general reactions are shown in Figure 12-6.

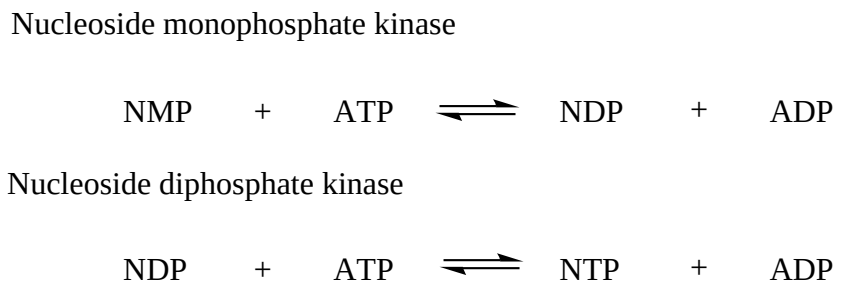


Figure 12-6: Two of the reactions catalyzed by the kinase enzymes.

The formation of the deoxyribonucleoside triphosphates, dNTP, follows two different paths. In one path, a multienzyme system converts the appropriate nucleoside diphosphate to the corresponding deoxyribonucleoside diphosphate. Then nucleoside diphosphate kinase catalyzes the formation of the deoxyribonucleoside triphosphate. The other path occurs in certain microorganisms where there is a direct conversion of NTP to dNTP.

As easy as 1, 2, 3: AMP, ADP, and ATP

It is possible to hydrolyze ATP either to ADP plus phosphate, P_i, or to AMP plus pyrophosphate, PP_i. (The pyrophosphate will undergo further hydrolysis to two phosphates, 2 P_i.) ADP and P_i are the immediate precursors for the reformation of ATP. To produce ATP starting with AMP utilizes the enzyme adenylate kinase. This enzyme catalyzes the transfer of a phosphate group from an ATP to an ADP. This reaction results in the formation of two ADP molecules. (Adenylate kinase also catalyzes the reverse reaction.)

The easy transfer of phosphate groups between nucleotides creates a metabolic network for the transfer of energy. The key to this network is the intercellular production of ATP.

Where It All Comes From

One of the purposes of the food we eat, of course, is to supply energy, with carbohydrates and fats being the major sources of energy. Digestion breaks polysaccharides into glucose and other monosaccharides, whereas fats are broken into glycerol and fatty acids. Catabolism converts these energy sources primarily to ATP. Proteins are broken into amino acids, which usually do not serve as energy sources. (We explain the details of these reactions later in this book.) Glucose produces 36 ATP molecules. This is an average of 6 ATPs per carbon. The step-by-step energy change for glucose is in Table 12-3. Other carbohydrates give a similar yield.

Table 12-3 ATP Yield for Each Step in the Metabolism of Glucose

Chemical Steps	*Number of ATP Molecules Produced*
Activation (conversion of glucose to 1,6-fructose diphosphate)	-2
Oxidative phosphorylation 2(glyceraldehyde 3-phosphate $\rightarrow$ 1,3-diphosphoglycerate), producing 2 NADH + H^+ in cytosol	4
Dephosphorylation 2(1,3-diphosphoglycerate $\rightarrow$ pyruvate)	4
Oxidatie decarboxylation 2(pyruvate $\rightarrow$ acetyl CoA), producing 2 NADH + H^+ in mitochondrian	6
Oxidation of two C_2 fragments in citric acid and oxidative phosphorylation common pathway, producing 12 ATP for each C_2 fragment	24
Total	36

Each fat molecule hydrolyzes to a glycerol and three fatty acid molecules. Glycerol produces 20 ATPs per molecule. The energy production from a fatty acid will vary with the identity of the particular acid. Stearic acid, $C_{18}H_{36}O_2$, produces a total of 146 ATPs per molecule. This amounts to an average of 8.1 ATPs per carbon. The step-by-step energy change for stearic acid is shown in Table 12-4. Other fatty acids give a similar yield.

Table 12-4 ATP Yield for Each Step in the Metabolism of Stearic Acid

Chemical Steps	*Happens*	*ATP Molecules Produced*
Activation (stearic acid → stearyl CoA)	Once	-2
Dehydrogenation (acetyl CoA → transenoyl CoA), producing $FADH_2$	8 times	16
Dehydrogenation (hydroxyacyl CoA → keto acyl CoA), producing $NADH + H^+$	8 times	24
C_2 fragment (acetyl CoA → common catabolic pathway), producing 12 ATP per C_2 fragment	9 times	108
Total	146	

What happens if you stop eating?

Starvation is the total deprivation of food. Here is what happens during starvation: Initially, the body utilizes its glycogen reserves. Then it moves on to its fat reserves — the first ones are those around the heart and kidneys. Finally, the body relies on the reserves found in the bone marrow. Early in a total fast, the body metabolizes protein at a rapid rate. The amino acids are converted to glucose, because the brain prefers glucose. These proteins come from the skeletal muscles, blood plasma, and other sources in a process that produces a quantity of nitrogen-containing products, which need to be excreted. Excretion requires large quantities of water, and the resulting loss of water may lead to death by dehydration. If the starvation continues, the brain chemistry adjusts to accept fatty acid metabolites, which uses the last of the fat reserves. Finally, the body resorts to structural proteins, systems begin to fail rapidly, and death follows quickly.

Chapter 13

ATP: The Body's Monetary System

In This Chapter

- Checking out carbohydrate metabolism and examining the citric acid cycle
- Finding out about electron transport and oxidative phosphorylation
- Seeing how biosynthesis takes place

Here we examine a number of general processes that either produce or consume energy. Breaking down molecules often produces energy. The breakdown of one molecule is often coupled with the synthesis of another, and this other synthesized molecule is often adenosine triphosphate, or ATP. *Catabolism* is the breaking down of molecules to provide energy. *Anabolism* is the building of molecules. These two processes combine to give metabolism. *Metabolism* comprises all reactions in biological systems.

As you can see in Chapter 12, the "currency" in biological systems is ATP. There are other energy-containing molecules, but the rate of exchange to ATP is the reference. The breakdown of certain molecules produces the currency of ATP, and there is a cost involved in the synthesis of other molecules. Polysaccharides and fats are like "banks" that store energy for later use.

Metabolism 1: Glycolysis

The *Embden-Meyerhof pathway,* or *glycolysis,* is a primitive means of extracting energy from organic molecules. The process converts glucose to two lactic acid molecules in an anaerobic (without oxygen) process. Nearly all forms of life, whether a person or a jellyfish, utilize glycolysis. All carbohydrates follow this pathway. Aerobic (utilizing oxygen) processing of carbohydrates uses pyruvate derived from glycolysis. (Alcoholic fermentation also produces pyruvate from glucose. The glucose is converted to two ethanol molecules and two CO_2 molecules.) Glycolysis is a two-part process, which we label Phase I and Phase II. Figures 13-1 and 13-2 help illustrate the upcoming, ahem, rather *involved* discussion. You may want to refer back to these figures as you read.

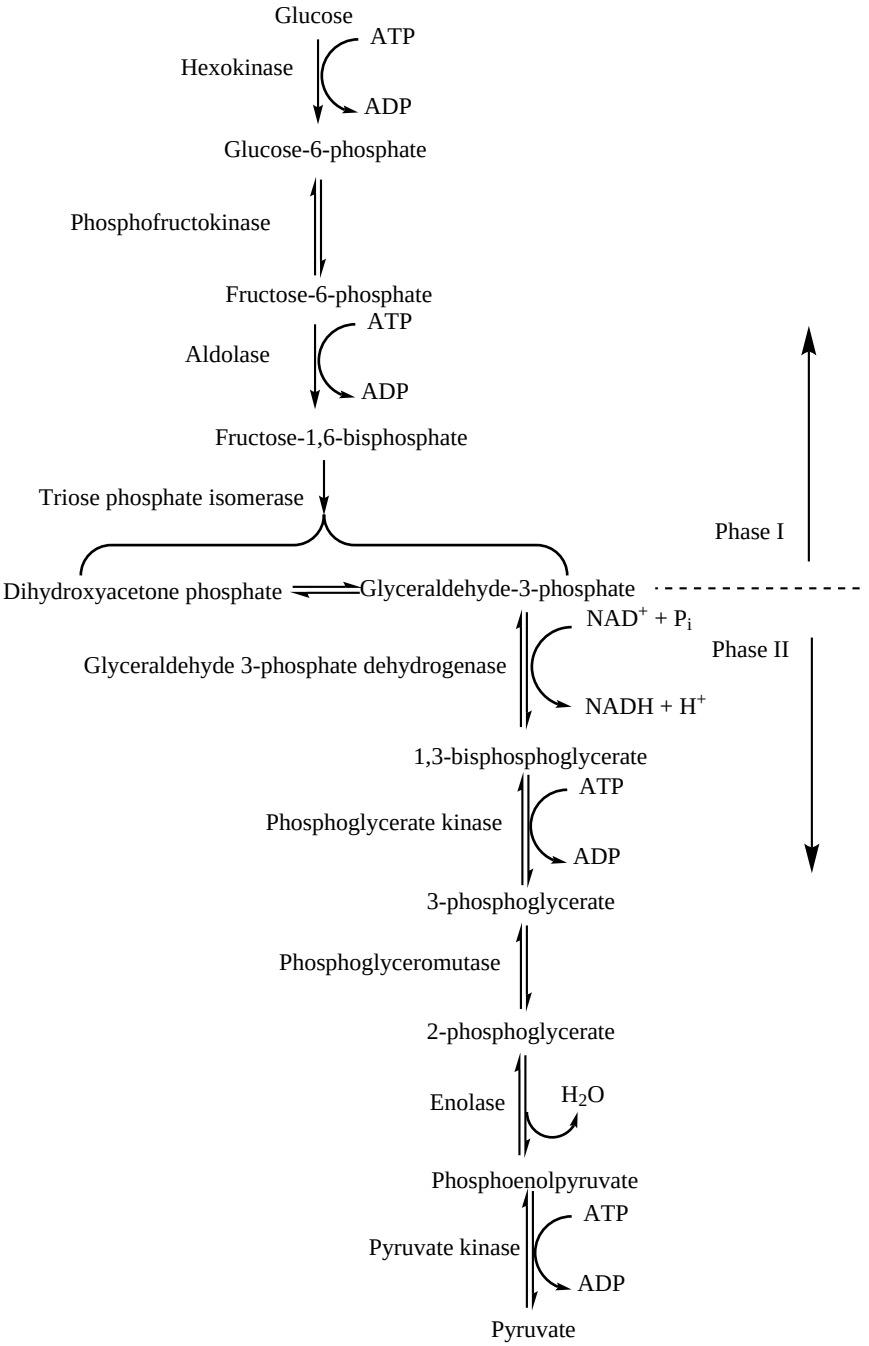


Figure 13-1: Steps in glycolysis.

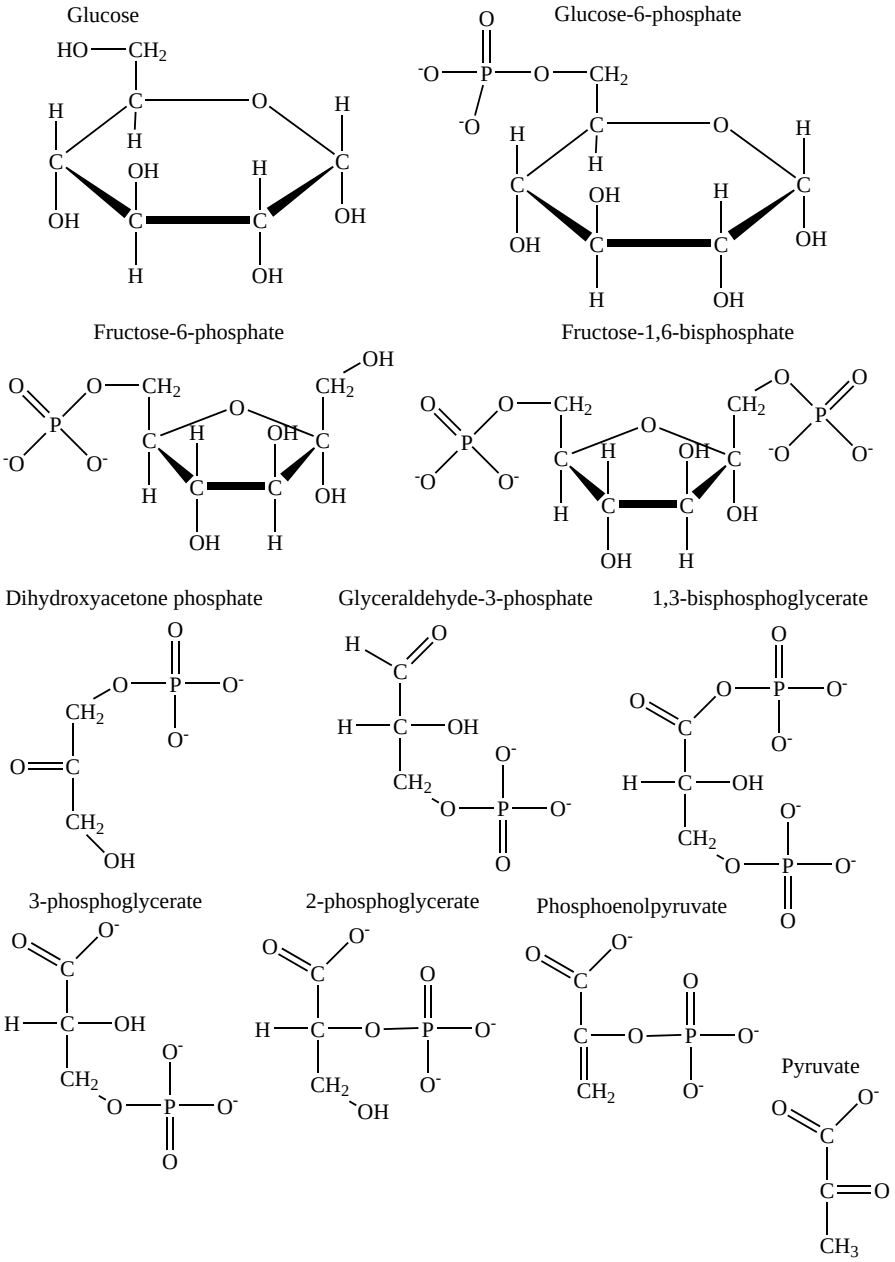


Figure 13-2: Molecules involved in glycolysis.

Glucose: Where it all starts

As we mentioned, glycolysis occurs in two phases: Phase I and Phase II.

Phase 1

As glucose enters the cell, it undergoes immediate phosphorylation to glucose-6-phosphate — the first step in Phase I. The phosphate comes from ATP, and the enzyme hexokinase, with the aid of Mg^{2+}, catalyzes the transfer. Thus, the first step in the production of energy requires an investment of energy, which is necessary to activate the glucose in a reaction that is not easy to reverse. In addition, the presence of the charged phosphate group makes it difficult for this and other intermediates to diffuse out of the cell.

The enzyme phosphoglucose isomerase then catalyzes the isomerization of glucose-6-phosphate to fructose-6-phosphate. This results in a compound with a primary alcohol group, which is easier to phosphorylate than the hemiacetal originally present. Fructose-6-phosphate then reacts with another molecule of ATP to form fructose-1,6-bisphosphate. The enzyme for this step is phosphofructokinase — (try saying that ten times fast!) — and this enzyme requires Mg^{2+} to be active. This is the major regulatory step in glycolysis. ATP inhibits this enzyme, whereas AMP activates it.

Aldolase enzymatically cleaves the fructose-1,6-bisphosphate into two triose phosphates. These triose phosphates are dihydroxyacetone phosphate and glyceraldehyde-3-phosphate. The dihydroxyacetone phosphate isomerizes to glyceraldehyde-3-phosphate to complete Phase I. Triose phosphate isomerase catalyzes this isomerization. (You see why we suggested following along with Figures 13-1 and 13-2?)

The net result of Phase I is the formation of two molecules of glyceraldehyde-3-phosphate, which costs two ATP molecules, and produces no energy.

Phase 11

Phase II begins with the simultaneous phosphorylation and oxidation of glyceraldehyde-3-phosphate to form 1,3-bisphosphoglycerate. Glyceraldehyde-3-phosphate dehydrogenase catalyzes this conversion. Inorganic phosphate is the source of the phosphate. NAD^+ is the coenzyme and oxidizing agent. NAD^+ reduces to NADH.

There is a high-energy acyl phosphate bond present in 1,3-bisphosphoglycerate. Phosphoglycerate kinase, in the presence of Mg^{2+}, catalyzes the direct transfer of phosphate from 1,3-bisphosphoglycerate to ADP. This results in the formation of ATP and 3-phosphoglycerate. Because the formation of ATP involves direct phosphate transfer, this process is called *substrate-level phosphorylation* to avoid confusion with *oxidative phosphorylation* (discussed later). Phosphoglyceromutase then catalyzes the transfer of a phosphate

group from C-2 to C-3, thus converting 3-phosphoglycerate to 2-phosphoglycerate. After that, dehydration occurs to form phosphoenolpyruvate (PEP), which contains a nigh-energy phosphate bond. The enzyme catalyzing the reaction is enolase.

The final, irreversible step is a second substrate-level phosphorylation. Here, an ADP molecule receives a phosphate group from the PEP. The enzyme pyruvate kinase is necessary for this step. This enzyme requires not Mg^{2+}, but also K^{+}. Pyruvate is the other product. Whew!

During Phase II, two molecules of glyceraldehyde-3-phosphate (from Phase I) form two molecules of pyruvate with the formation of four molecules of ATP and two molecules of NADH.

The pyruvate produced by glycolysis has several fates. When there is plenty of oxygen, the pyruvate enters the Krebs cycle, the electron transport chain, and oxidative phosphorylation pathways as Acetyl-CoA. This results in the production of more ATP and the total conversion to CO_2. If oxygen is lacking, vertebrates (you included) convert pyruvate to a related substance, lactate. Other organisms, such as yeast, convert pyruvate to ethanol and CO_2 — and that is why we have beer. These latter two possible fates yield less energy than the oxygen-rich fate.

Miles per gallon? Energy efficiency

Glycolysis is the initial conversion of carbohydrate to energy. After that there is the production of two ATP molecules, two NADH molecules, and two pyruvate molecules. The energy content of the ATP molecules is only 2 percent of the total energy present in each glucose molecule. This shows the relative inefficiency of anaerobic energy production. Fortunately, the pyruvate molecules will undergo further aerobic oxidation to increase this energy yield. The total energy output of anaerobic and aerobic oxidation of glucose is 30–32 ATP molecules, which accounts for about 30 percent of the total energy present in glucose. Much of the remaining energy is available as heat for warm-blooded animals.

Going in reverse: Gluconeogenesis

Gluconeogenesis is a series of reactions that generate glucose from non-carbohydrate sources. This pathway is necessary when the supply of carbohydrates is inadequate (something that is rare in our lives). The non-carbohydrate sources include lactate, pyruvate, some amino acids, and glycerol. In many ways, gluconeogenesis is the reverse of glycolysis. Figure 13-3 summarizes the steps of gluconeogenesis. (The formation of glucose in plants utilizes the process of photosynthesis.)

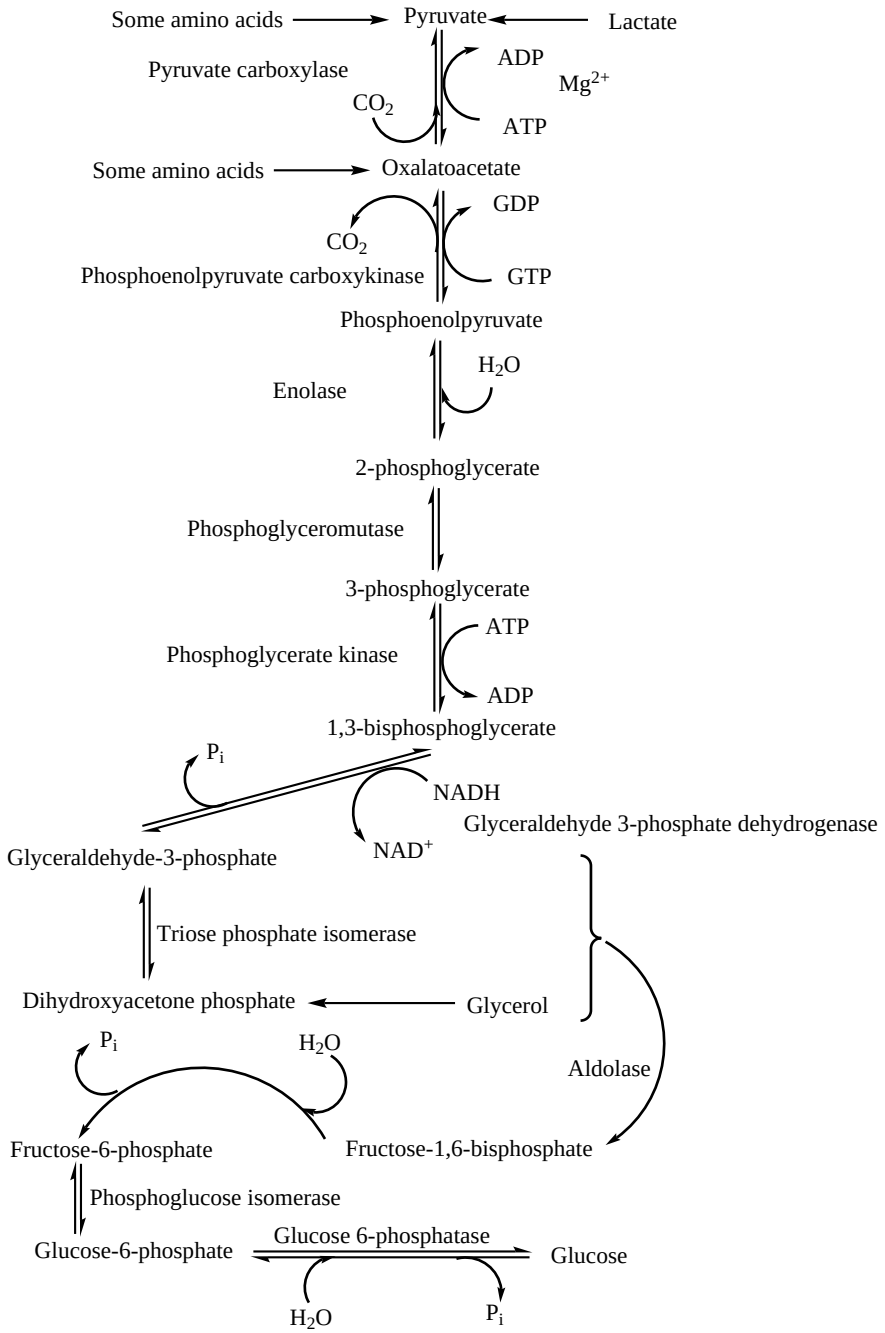


Figure 13-3: Steps in gluconeogenesis.

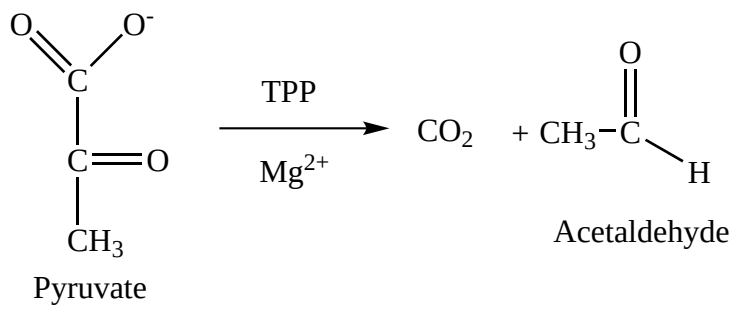
The presence of many of the same intermediates enables the use of many of the same enzymes in both glycolysis and gluconeogenesis. The differences (four enzymes) between the two systems allow regulation, so that the processes don't cancel each other. Regulation is also possible by isolating the two pathways in different organs. Other carbohydrates may also form.

Alcoholic fermentation: We'll drink to that

Under anaerobic conditions, yeast and other organisms convert pyruvate to ethanol and carbon dioxide. This process is accompanied by the oxidation of NADH to NAD^+. The NAD^+ is used in glycolysis. During this process, there is a net generation of two ATP molecules.

The first step in alcoholic fermentation is the decarboxylation of pyruvate to carbon dioxide and acetaldehyde. The enzyme pyruvate decarboxylase, along with the cofactors Mg^{2+} and TPP (thiamin pyrophosphate), catalyze this step. The enzyme alcohol dehydrogenase, along with the coenzyme NADH, catalyzes the conversion of acetaldehyde to ethanol. Makes you really appreciate that shot of tequila, doesn't it? Figure 13-4 summarizes these steps.

1. Pyruvate decarboxylase reaction



2. Alcohol dehydrogenase reaction

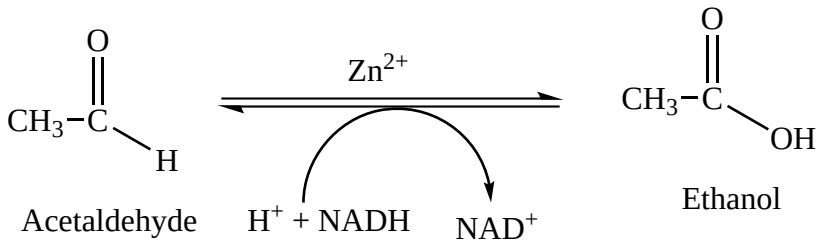


Figure 13-4: Steps in alcoholic fermentation.

Metabolism II: Citric Acid (Krebs) Cycle

The *citric acid cycle* and *oxidative phosphorylation* are the aerobic processes of catabolism that produce energy (ATP). The citric acid cycle is also known as the *Krebs cycle* and also as the *tricarboxylic acid cycle* (TCA). The primary entry molecule for this series of reactions is acetyl-CoA (short for acetyl-coenzyme A). The sources of acetyl-CoA are pyruvate from glycolysis, certain amino acids, or the fatty acids present in fats. The structure of acetyl-CoA is shown in Figure 13-5. Note: these processes take place in the *mitochondria,* the energy factories of the cell.

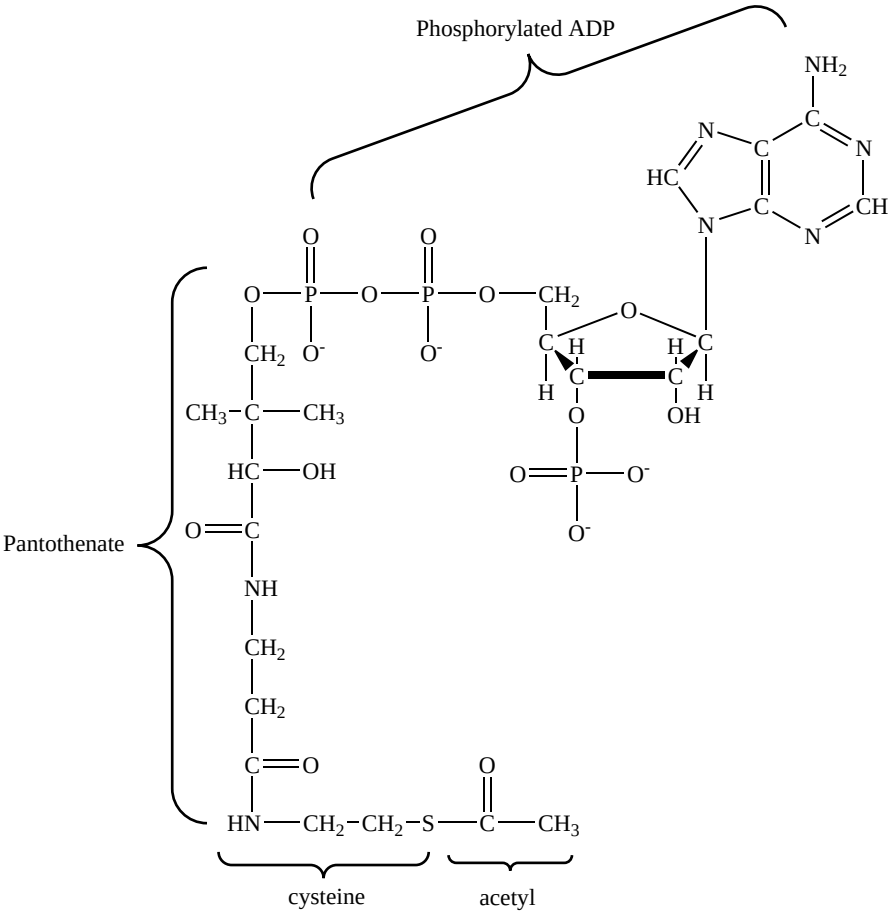


Figure 13-5: Structure of acetyl-CoA.

In addition to being an energy source, acetyl-CoA is the starting material for the synthesis of a number of biomolecules. In the next few sections, we discuss the citric acid cycle. The general cycle is shown in Figure 13-6, and the structures are shown in Figure 13-7.

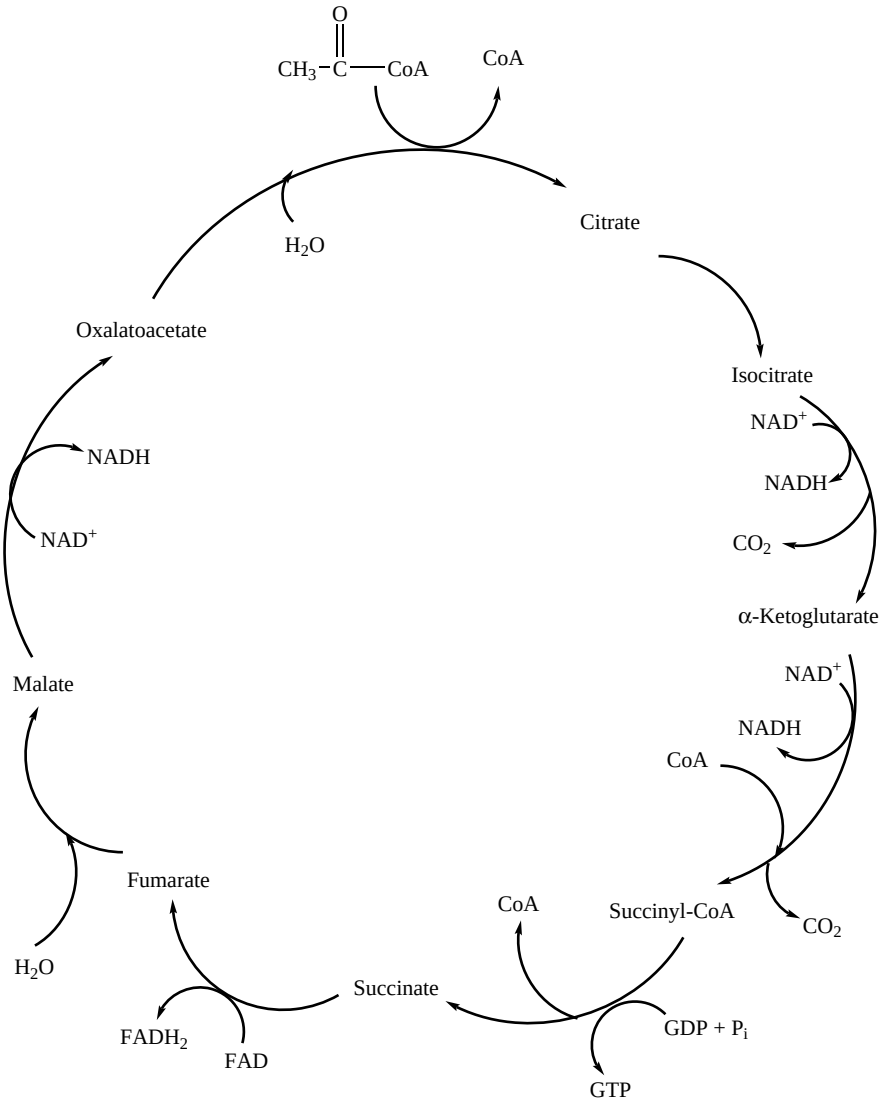
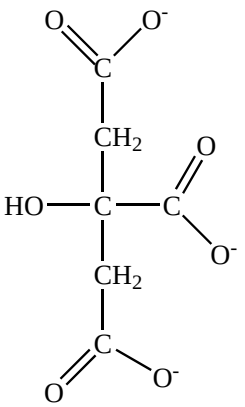
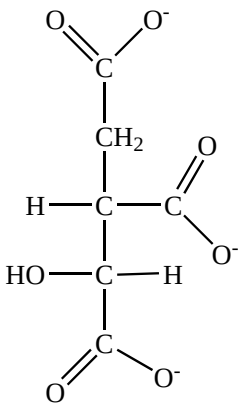


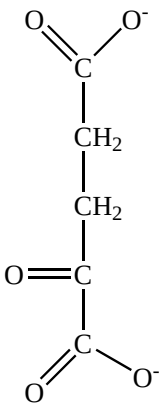
Figure 13-6: Citric acid (Krebs) cycle.



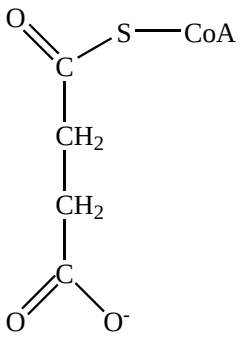
Citrate



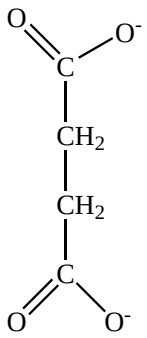
Isocitrate



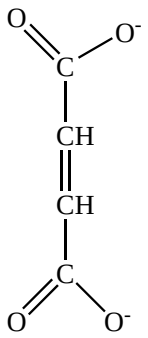
α-Ketoglutarate



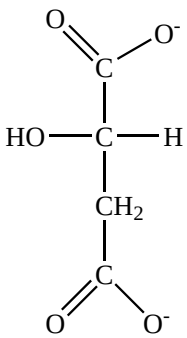
Succinyl-CoA



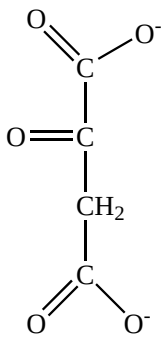
Succinate



Fumarate



Malate



Oxalatoacetate

Figure 13-7: Structures of molecules involved in the citric acid (Krebs) cycle.

Let's get started: Synthesis of acetyl-CoA

The synthesis of acetyl-CoA is a multi-step process. Figure 13-8 shows a simplified version of this process.

These steps are coupled to preserve the free energy produced by the decarboxylation. In the first step, pyruvate combines with TPP (thiamin pyrophosphate) and undergoes decarboxylation. The pyruvate dehydrogenase component of the multi-enzyme complex catalyzes this step. During the second step, the TPP undergoes oxidation, which yields an acetyl group (refer back to Figure 13-8). This acetyl group transfers to lipomide. In this reaction, the oxidant is the disulfide group of lipomide, and acetyllipoamide forms in this step. The pyruvate dehydrogenase component also catalyzes this reaction. In the final step, the acetyl group of acetyllipoamide transfer to CoA to form acetyl CoA. The catalyst for this reaction is dihydrolipoyl transacetylase.

However, the process does not end with the formation of acetyl CoA. It is necessary to regenerate the oxidized form of lopoamide. The enzyme dihydrolipoyl dehydrogenase catalyzes this step. The two electrons from the oxidation transfer to FAD and then to NAD^+. Some of the important intermediates in these steps are shown in Figure 13-9.

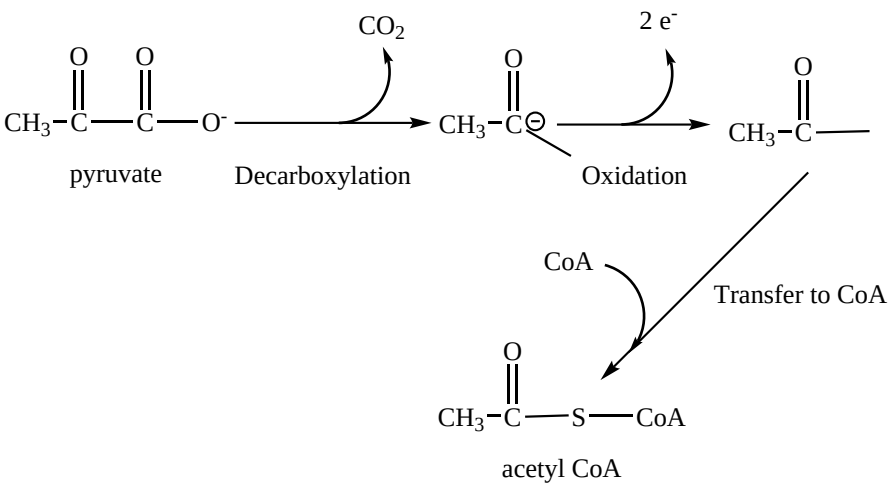
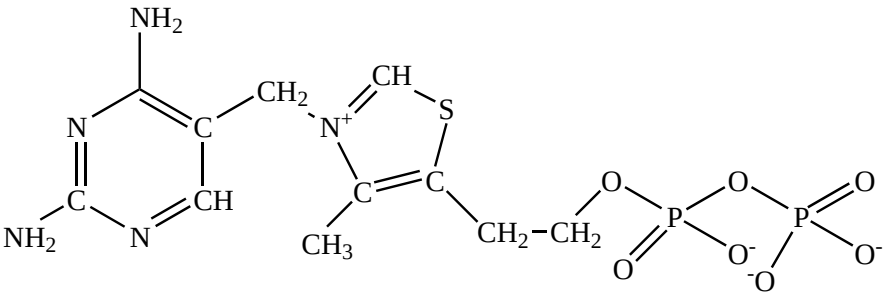
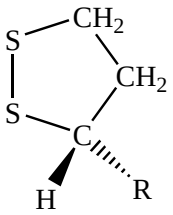


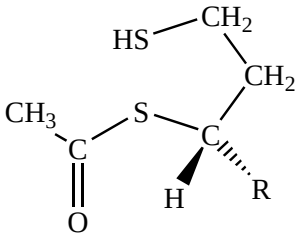
Figure 13-8: Simplified scheme for the formation of acetyl CoA.



Thiamine pyrophosphate (TPP)



Lipoamide



Acetyllipoamide

Figure 13-9: Structures of TPP, lipomide, and acetyllipo-amide.

Three's a crowd: Tricarboxylic acids

When acetyl-CoA enters the citric acid cycle, it interacts, in the presence of citrate synthase, with oxaloacetate. This interaction results in the transfer of the acetyl group to the oxaloacetate to form citrate. The hydrolysis of the thioester linkage of the acetyl-CoA releases a large amount of energy.

The enzyme aconitase, with Fe^{2+} as a cofactor, catalyzes the isomerization of citrate to isocitrate. For a time, cis-aconitate, derived aconitase, was thought to be a part of the citric acid cycle. However, even though the structure of cis-aconitate is related to the other tricarboxylic acids, it is *not* part of the citric acid cycle. The structure of cis-aconitate is in Figure 13-10.

Just a little gas: Oxidative decarboxylation

The next step is the conversion of isocitrate to α-ketoglutarate. The molecule passes through the intermediate oxalosuccinate. The isocitrate binds to the enzyme isocitrate dehydrogenase. During this process, the coenzyme NAD^+ undergoes reduction. Both ATP and NADH are negative factors in the allosteric regulation of isocitrate dehydrogenase, whereas ADP is a positive factor. This is an important mechanism to control the production of ATP.

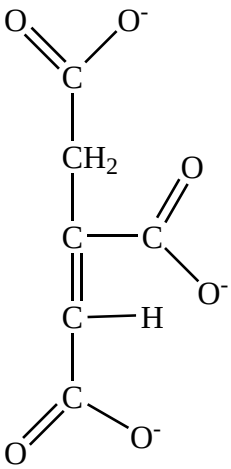


Figure 13-10: Structure of cis-aconitate.

Production of succinate and GTP

The conversion of α-ketoglutarate to succinate requires two steps. The α-ketoglutarate must bind to the enzyme to form an α-ketoglutarate dehydrogenase complex. This reaction requires the same cofactors as needed for the formation of acetyl-CoA. The result of this reaction is the elimination of carbon dioxide and the formation of succinyl-CoA. This process is irreversible under physiological conditions.

In the second step, succinyl-CoA separates to form succinate and release energy, which is harnessed by the conversion of GDP to GTP. This substrate-level phosphorylation is catalyzed by succinyl-CoA synthetase. (GTP contains about the same energy as ATP and can substitute for ATP.)

Oxaloacetate regeneration

The regeneration of oxaloacetate completes the cycle, requiring three reactions which, together, convert a methylene to a carbonyl group. First, a hydrogen atom is removed from each of two adjacent carbon atoms, resulting in the formation of a double bond. Next, a water molecule adds to the double bond. Finally, the removal of two hydrogen atoms yields the appropriate α-keto group. Succinate dehydrogenase catalyzes the first of these reactions. The prosthetic group, FAD, accepts the two hydrogen atoms by covalently binding to the enzyme. Fumarase catalyzes the next step. The final oxidation utilizes the enzyme malate dehydrogenase with the coenzyme NAD^+. The oxaloacetate is now ready to begin the cycle again.

Amino acids as energy sources

Although carbohydrates are the most readily available energy source, there are situations where amino acids can serve as energy sources. This is important for carnivores (like ourselves), who live on a high protein diet. The utilization of amino acids as energy sources is also important during hypoglycemia, fasting, and starvation.

The process begins with the removal of the amino group. This usually occurs through *transamination,* which is the transfer of an amino group from one molecule to another. Any amino acid other than threonine, proline, and lysine will undergo this process. Usually, the amino group transfers to the keto carbon of α-ketoglutarate, oxalatoacetate, or pyruvate to form glutamate, aspartate, or alanine, respectively. Specific transaminases are necessary and the coenzyme pyridoxal phosphate catalyzes this process. A second transamination is involved in the process of transforming aspartate and alanine to glutamate.

Oxidative deamination converts glutamate to α-ketoglutarate. This process, which occurs primarily in the liver, releases an ammonium ion. The reverse reaction, glutamate synthesis, is one of the few reactions that occurs in animals in which inorganic nitrogen is converted into organic nitrogen. The ammonium ion resulting from oxidative deamination may enter one or more biosynthetic pathways or the urea cycle. Most vertebrates convert the ammonium ion to urea, which is excreted in the urine. Most marine organisms, including fish, eliminate ammonia directly, whereas birds, insects, and reptiles convert the ammonium ion to uric acid.

The products of transamination, oxidative deamination, and further modification of the remaining portion of the amino acid produce one of the intermediates in glycolysis or the citric acid cycle. This is the fate of all the amino acids — some of the amino acids go through one intermediate, whereas others require more intermediates. Figure 13-11 shows where each of the amino acids enters glycolysis or the citric acid (Krebs) cycle. Some of the amino acids have more than one entry point.

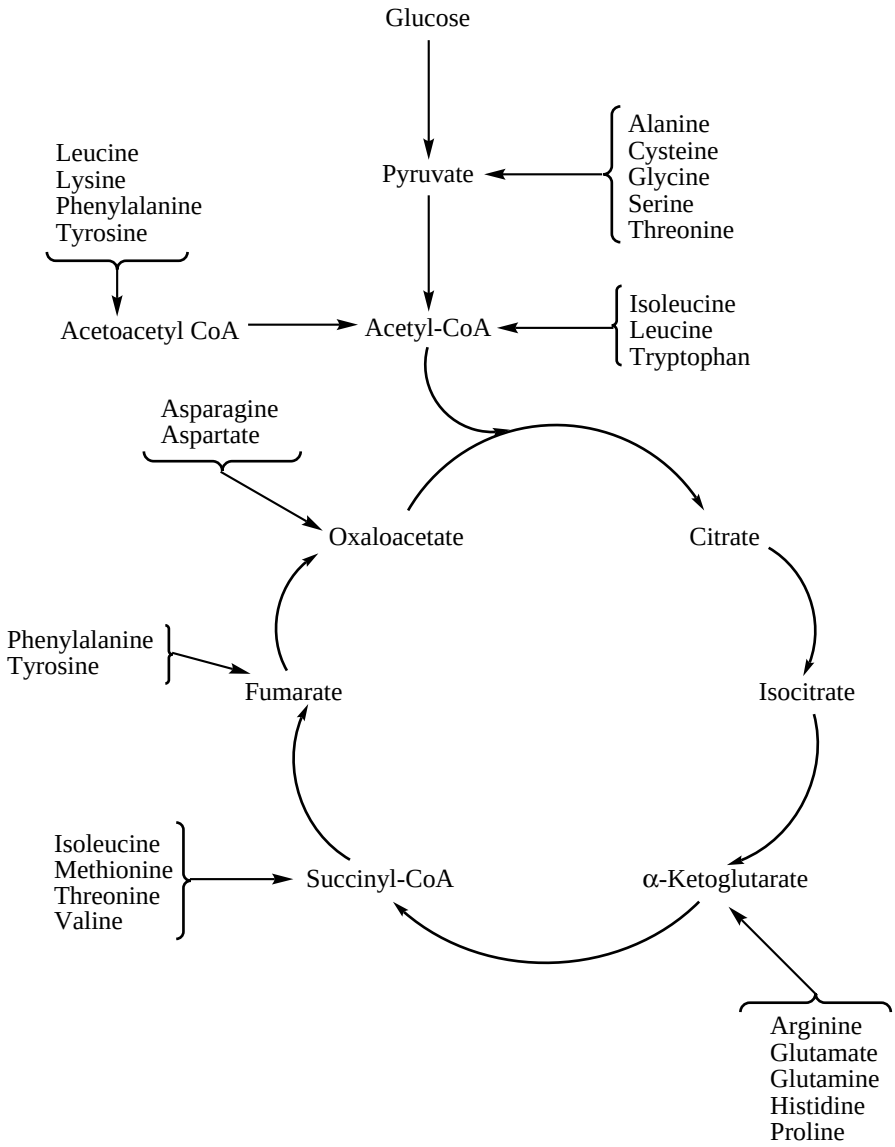


Figure 13-11: Fate of the amino acids.

Electron Transport and Oxidative Phosphorylation

The production of NADH and $FADH_2$ by the citric acid cycle supplies the materials for the next phase: oxidative phosphorylation. These reduced coenzymes transport the electrons derived from the oxidation of pyruvate. The final fate of these electrons is the reduction of oxygen to water.

The details of oxidation phosphorylation are not as easy to study as glycolysis and the citric acid cycle because the processes take place within the mitochondria, where many of the proteins involved are integrated into the walls. In addition, many of the processes are coupled. The separate components of a *coupled* process must not only be in close proximity, but often need to be in a specific arrangement.

The electron transport system

A number of species in the mitochondria must undergo oxidation-reduction reactions. Oxidation involves a loss of electrons, whereas reduction involves a gain of electrons. These processes are coupled in that the electrons lost must equal the electrons gained. The reduction potential indicates how easily a molecule undergoes oxidation or reduction. The molecular players that are important to the electron transport system are the pyridine-linked dehydrogenases, flavin-linked dehydrogenases, iron-sulfur proteins, ubiquinones, and cytochromes.

Off on a tangent: Dealing with reduction potentials

The standard for reduction potentials is the reaction:

$$2\ H^+(aq) + 2\ e^- \quad H_2(g)$$

Under standard conditions (25°C, P_{H2} = 1 atm, and $[H^+]$ = 1.0 M), the standard reduction potential is $E^\circ = 0.00$ V. Under physiological conditions in humans the value is –0.42 V (designated as E'°), because the conditions are not standard.

Table 13-1 lists a number of physiological reduction potentials. We show you how to use these entries later. The values in the table are arranged in order of increasing potential. The higher the value, the better the reaction is at oxidation, and the lower the value, the better the reaction is at reduction.

Table 13-1 Some Physiological Reduction Potentials (E'°)

	E'°(volts)
Ferredoxin-Fe^{3+} + e^- ⇌ Ferredoxin-Fe^{2+}	–0.43
2 H^+(aq) + 2 e^- ⇌ H_2(g)	–0.42
α-Ketoglutarate + CO_2 + 2 H^+ + 2 e^- ⇌ Isocitrate	–0.38
NAD^+ + H^+ + 2 e^- ⇌ NADH	–0.32
FAD + 2 H^+ + 2 e^- ⇌ $FADH_2$	–0.22
Riboflavin + 2 H^+ + 2 e^- ⇌ Riboflavin-H_2	–0.20
Dihydroxyacetone phosphate + 2 H^+ + 2 e^- ⇌ Glycerol 3-phosphate	–0.19
Pyruvate + 2 H^+ + 2 e^- ⇌ Lactate	–0.19
Oxaloacetate + 2 H^+ + 2 e^- ⇌ L-Malate	–0.17
Fumarate + 2 H^+ + 2 e^- ⇌ Succinate	+0.03
Cytochrome b-Fe^{3+} + e^- ⇌ Cytochrome b-Fe^{2+}	+0.08
Cytochrome c-Fe^{3+} + e^- ⇌ Cytochrome c-Fe^{2+}	+0.22
Cytochrome c_1-Fe^{3+} + e^- ⇌ Cytochrome c_1-Fe^{2+}	+0.23
Cytochrome a-Fe^{3+} + e^- ⇌ Cytochrome a-Fe^{2+}	+0.29
Cytochrome a_3-Fe^{3+} + e^- ⇌ Cytochrome a_3-Fe^{2+}	+0.38
1/2 O_2 + 2 H^+ + 2 e^- ⇌ H_2O	+0.82

Each reaction in Table 13-1 is known as a *half-reaction.* It takes two half-reactions — one oxidation and one reduction — to produce a complete (oxidation-reduction) reaction. The electrons lost (oxidation) must equal the electrons gained (reduction). For this reason, electrons only appear in the half-reaction, but never in the overall reaction.

By convention, the reactions in Table 13-1 all appear as *reduction* half-reactions. To convert any of these to an *oxidation* half-reaction, you must do two things. First, reverse the reaction, and then reverse the sign of E'°. In an oxidation-reduction reaction, the overall reaction is created by combining (adding) an oxidation reaction with a reduction reaction. Before adding the two reactions, though, make sure that the electrons in each reaction are equal. This may require multiplying one or both of the reactions by a value to make sure the

electrons are equal. (Multiply the reactions only — do not change the value of E'° [other than a sign change].) For example, look at the following reactions from the table:

$NAD^+ + 2\ e^-$ NADH –0.32

Cytochrome b-Fe^{3+} + e^- Cytochrome b-Fe^{2+} +0.075

Let's now change the first reaction to an oxidation:

NADH $NAD^+ + H^+ + 2\ e^-$ +0.32

If we now want to combine these reactions, we need to multiply the cytochrome reaction by two (so both reactions now involve two electrons):

2 Cytochrome b-Fe^{3+} + 2 e^- 2 Cytochrome b-Fe^{2+} +0.08

The number of electrons lost must equal to electrons gained. Also, notice that only the reaction is doubled, not the voltage. We can now combine these two reactions, canceling the electrons from both sides:

NADH $NAD^+ + H^+ + 2\ e^-$ +0.32 V

2 Cytochrome b-Fe^{3+} + 2 e^- 2 Cytochrome b-Fe^{2+} +0.08 V

NADH + H^+ + 2 Cytochrome b-Fe^{3+} 2Cytochrome b-Fe^{2+} + NAD^+ + $2H^+$ +0.40 V

The final reaction will have no electrons. Other species may cancel, if they appear on both sides of the reaction arrow. Any time the sum of the two potentials is positive, the reaction produces energy. Conversely, a negative value means the reaction requires energy. The greater the value of the sum, the greater the amount of energy produced.

Pyridine-linked dehydrogenases

In order for these enzymes to function, the coenzymes NAD^+ or $NADP^+$ are necessary. The coenzymes may be in either the oxidized or the reduced forms. If the general form of the substrate in the reduced form is $Z\text{-}H_2$, and in the oxidized form, it is Z, then the reaction will be:

$Z\text{-}H_2$ + NAD^+ (or $NADP^+$) Z + NADH (or NADPH) + H^+

There are more than 200 pyridine-linked dehydrogenases. The majority of NAD^+-linked dehydrogenases are involved in aerobic respiration. Most of the $NADP^+$-linked dehydrogenases are involved in biosynthesis.

Flavin-linked dehydrogenases

Enzymes (E) of this type require FAD or FMN as tightly bound prosthetic groups or coenzymes. Again, the species may be in either the oxidized or the reduced forms. The general reactions of this type are:

$$Z\text{-}H_2 + E\text{-}FAD \quad Z + E\text{-}FADH_2$$

$$Z\text{-}H_2 + E\text{-}FMN \quad Z + E\text{-}FMNH_2$$

NADH dehydrogenase, which contains the prosthetic group FMN, is the enzyme responsible for transporting electrons from NADH to the next acceptor in the electrons transport chain. There are other flavin-linked dehydrogenases — for example, succinate dehydrogenase.

Iron-sulfur proteins

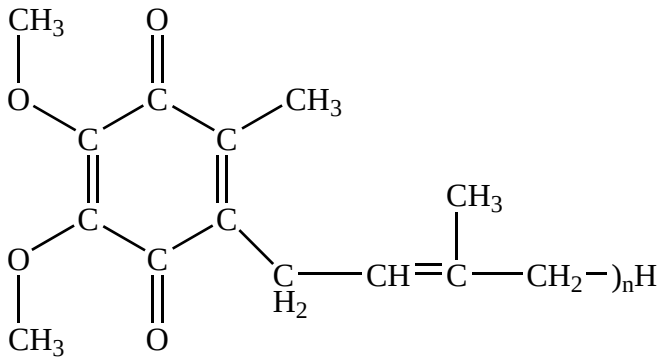
The chief characteristics of iron-sulphur proteins are the presence of iron and sulfur, as S^{2-}. The electron transporting ability of these proteins is the Fe^{2+}/Fe^{3+} couple. Several of these proteins are associated with the electron transport chain, where they are complexed to other respiratory species. Examples include succinate dehydrogenase, with two iron-sulfur centers, and NADH dehydrogenase, with four iron-sulfur centers.

Ubiquinones

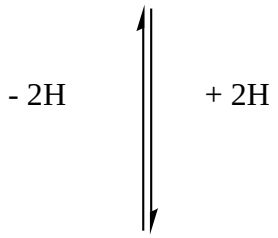
The *ubiquinones* are a group of coenzymes that are fat-soluble. Coenzyme Q (CoQ) is an example of an ubiquinone. The oxidation-reduction center is a derivation of quinine, and the fat-solubility is enhanced by the presence of a long hydrocarbon chain, containing a series of isoprene units. Many of the different ubiquinones differ only in the number of isoprene units present. The oxidized form of coenzyme Q is simply CoQ whereas the reduced form is $CoQH_2$. The general structures of both the oxidized and reduced forms of a ubiquinone appear in Figure 13-12.

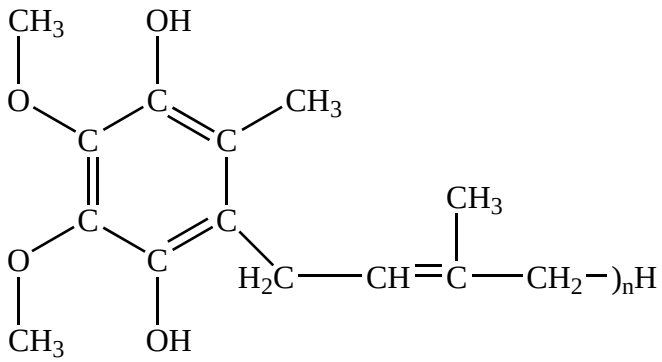
Cytochromes

The *cytochromes* are a group of proteins containing a heme group. Like the iron-sulfur proteins, the oxidation-reduction couple is Fe^{2+}/Fe^{3+}. The three general classes of cytochromes are a, b, and c. The derivation of the class names relates to spectral studies done during the first isolation of these molecules. Cytochromes occur in both the mitochondria and the endoplasmic reticulum. The heme group, present in all cytochromes, is like the heme groups present in myoglobin and hemoglobin. In all cases, the central portion of the group is identical; differences derive from the attachment of side-chains to the heme core. Figure 13-13 shows the heme core and where the side-chains normally attach.



Oxidized ubiquinone





Reduced ubiquinone

Figure 13-12: General structures of the oxidized and reduced forms of a ubiquinone.

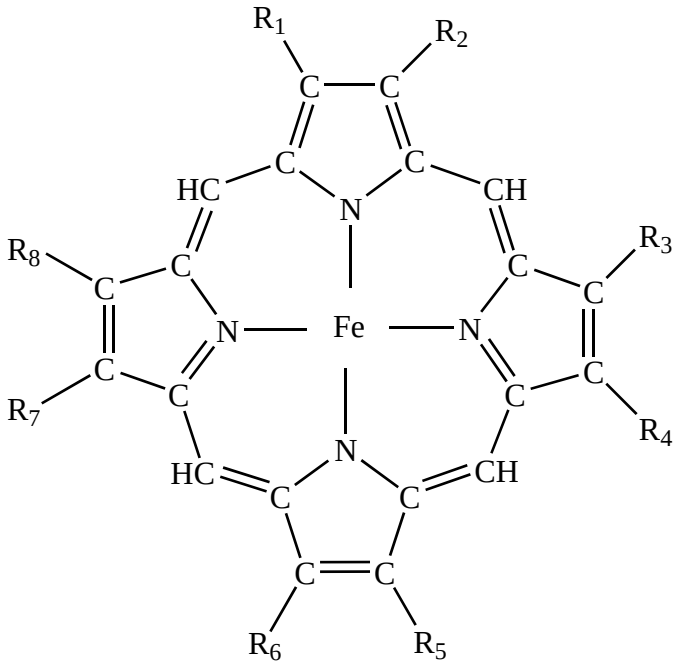


Figure 13-13: The heme core and attachment sites (R).

Five cytochromes (a, a_3, b, c, and c_1) have been identified as part of the electron transport chain of mammals. Cytochrome c, or *cyt c,* is easy to extract from cells, and therefore it is the most studied of the cytochromes. The structure of cytochrome c from different species is important to the study of biochemical evolution. Cytochromes a and a_3, cyt aa_3, occur together as a complex containing not only the expected two heme groups, but also two copper ions. The copper ions are part of another oxidation-reduction couple (Cu^{+}/Cu^{2+}). This complex, known as *cytochrome oxidase,* is the terminal cytochrome, which transfers electrons to O_2.

Interpersonal relationships

The members of the electron transport chain are grouped into four complexes with coenzyme Q (CoQ) and cytochrome c (cyt c) serving as links. One way of indicating the sequence of events in the electron transport chain appears in Figure 13-14. Figure 13-15 illustrates the same sequence emphasizing the cyclic nature of the steps. The processes take place in four complexes with linking CoQ and cytochrome c. These complexes are part of the inner mitochondrial membrane.

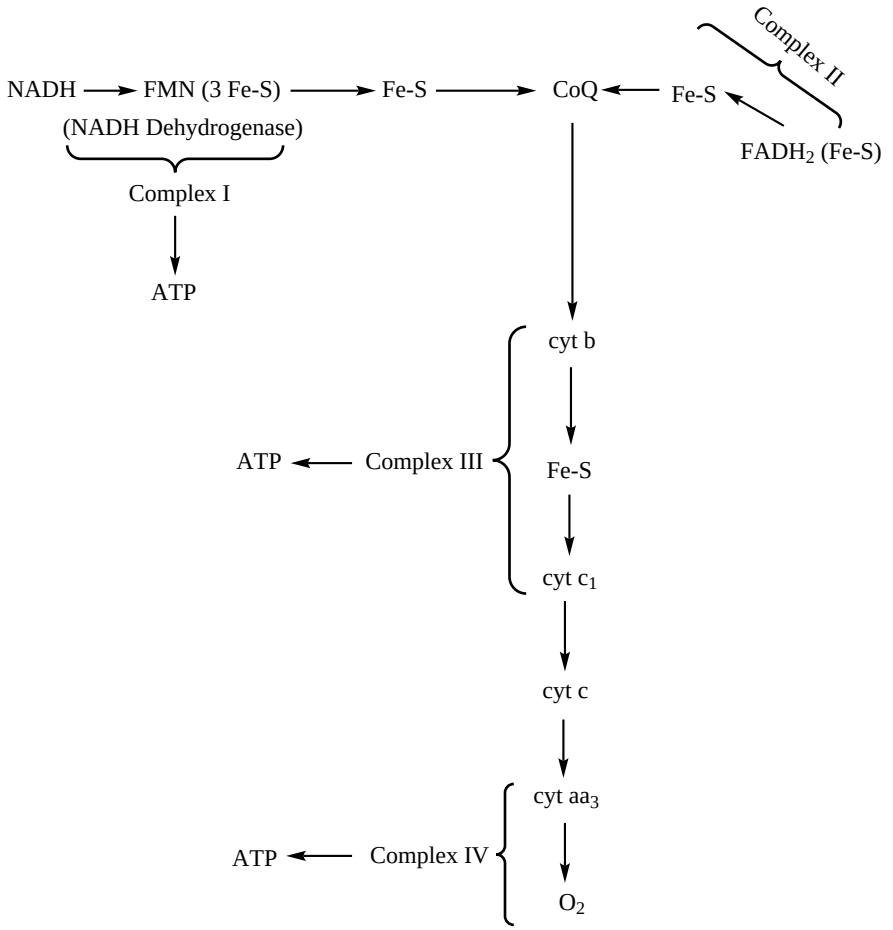


Figure 13-14: Steps in the electron transport chain.

Figure 13-15: Electron transport chain, emphasizing the cyclic nature of each of the processes.

The script: Oxidative phosphorylation

The processes of oxidative phosphorylation and the electron transport chain are closely coupled. It is only possible to oxidize the reduced forms of the coenzymes $FADH_2$ and NADH in the presence of ADP. The oxidations couple with the ADP transforming to ATP (phosphorylation).

If we calculate the oxidation-reduction potentials for NADH and $FADH_2$ reducing oxygen, we find:

$NADH \quad NAD^+ + H^+ + 2\ e^-$	+0.32 V
$\underline{1/2\ O_2 + 2\ H^+ + 2\ e^- \quad H_2O}$	$\underline{+0.82\ V}$
$1/2\ O_2 + H^+ + NADH \quad H_2O + NAD^+$	+1.14 V

And:

$FADH_2 \quad FAD + 2\ H^+ + 2\ e^-$	+0.22 V
$\underline{1/2\ O_2 + 2\ H^+ + 2\ e^- \quad H_2O}$	$\underline{+0.82\ V}$
$1/2\ O_2 + FADH_2 \quad H_2O + FAD$	+1.04 V

In both cases, the combination of the potentials is positive. Positive potentials refer to spontaneous processes, and spontaneous processes produce energy. Each NADH is capable of supplying sufficient energy to produce 2.5 ATP, and each $FADH_2$ can produce 1.5 ATP.

The play: Proposed mechanisms

The current proposed mechanism for oxidative phosphorylation is the *chemiosmotic hypothesis.* This hypothesis assumes that the hydrogen ion gradient is a significant factor promoting the conversion of ADP to ATP. The processes occurring in the four complexes present in the inner mitochondrial membrane result in a net transfer of hydrogen ions across the membrane.

The hydrogen ion transfer results in an increase in the hydrogen ion concentration in the space between the inner and outer mitochondrial membranes. It is necessary to move hydrogen ions back across the membrane. This transfer of hydrogen ions is necessary in the synthesis of ATP.

The box office: ATP production

The reactions from the anaerobic oxidation of glucose (glycolysis) and the aerobic oxidation of glucose result in the production of 32 molecules of ATP from every molecule of glucose. These reactions are:

Anaerobic:

$$\text{Glucose} + 2\ NAD^+ + 2\ ADP + 2\ P_i \rightarrow 2\ \text{Pyruvate} + 2\ NADH + 2\ H^+ + 2\ H_2O + 2\ ATP$$

Aerobic:

$$2\ \text{Pyruvate} + 5\ O_2 + 30\ ADP + 30\ P_i \rightarrow 6\ CO_2 + 34\ H_2O + 30\ ATP$$

Sum:

$$\text{Glucose} + 2\ NAD^+ + 5\ O_2 + 32\ ADP + 32\ P_i \rightarrow 6\ CO_2 + 36\ H_2O + 2\ NADH + 2\ H^+ + 32\ ATP$$

Involving the fats: β-oxidation cycle

Fatty acids may also serve as a source of ATP. Accomplishing this requires a series of reactions, known as β-*oxidation,* or the *fatty acid spiral,* to break down the fatty acid molecule. This series of reactions is a cyclic process. Some of the processes are oxidations, which require the coenzymes NAD^+ and FAD. This process also occurs in the mitochondria. The initiation of fatty acid oxidation requires activation of the relatively unreactive fatty acid molecule. The activated form is analogous to acetyl-CoA. In this case, the coenzyme A binds to the fatty acid to form a fatty acyl-CoA. Activation requires acyl-CoA synthetase and one molecule of ATP. The ATP uses two phosphates and becomes AMP.

At the inner mitochondrial membrane, the fatty acyl-CoA combines with the carrier molecule carnitine. Carnitine acyltransferase catalyzes this transfer. The fatty acyl-carnitine transports into the mitochondrial matrix, where it converts back to fatty acyl-CoA. With the mitochondrial matrix, a series of four reactions constitute the cycle known as β-oxidation. The name of this process refers to the oxidation of the second (β) followed by the loss of two carbons from the carboxyl end of the fatty acyl-CoA. Each trip around the cycle results in the removal of two carbon atoms, and the cycle continues until all the carbon atoms are removed. Figure 13-16 illustrates the general aspects of the cycle.

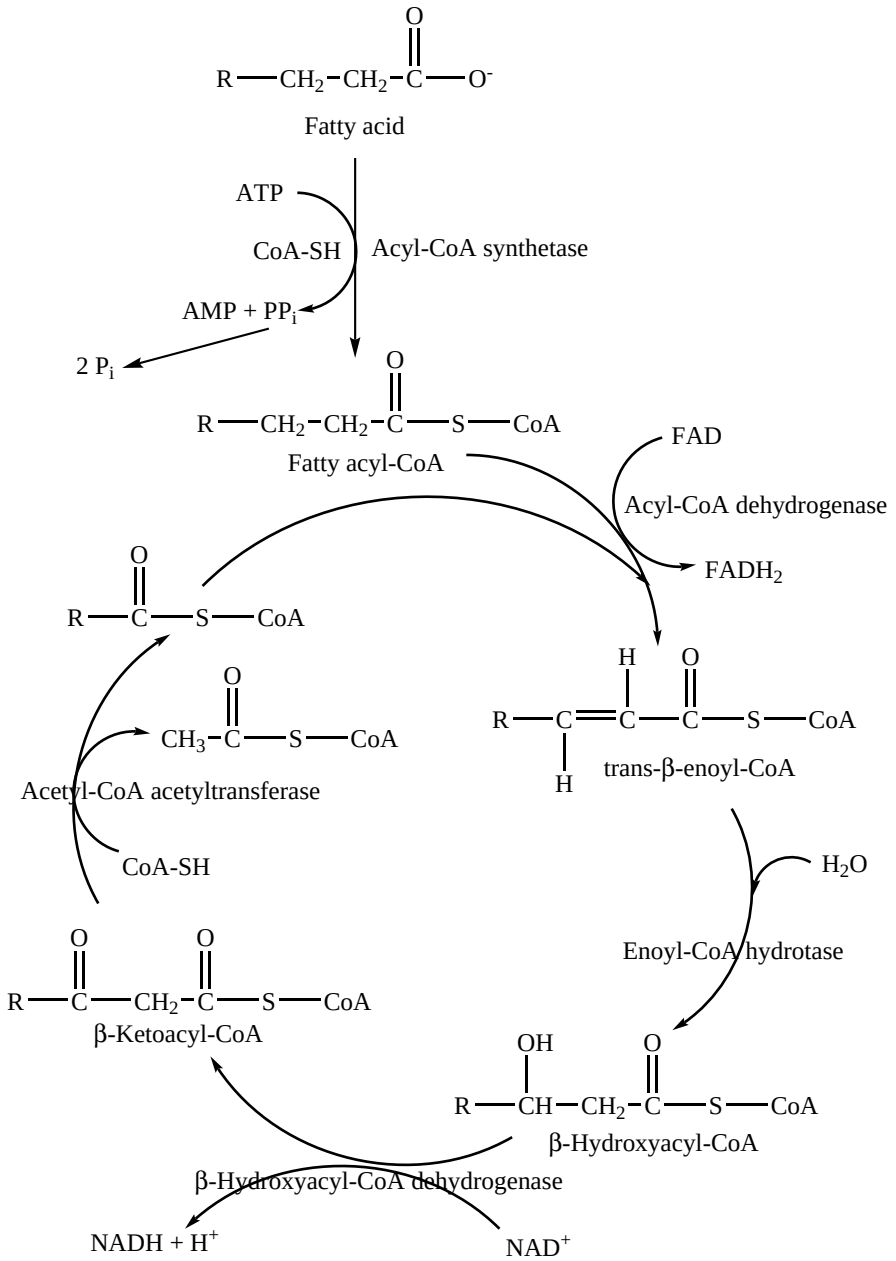


Figure 13-16: General steps in the β-oxidation cycle.

The first step in the cycle is an oxidation, with the catalyst being acyl-CoA dehydrogenase. During this step, coenzyme FAD accepts two hydrogen atoms. One of the hydrogen atoms is from the α carbon, and the other is from the β carbon atom. The process is stereospecific, producing the trans form. Elsewhere, the $FADH_2$ undergoes re-oxidation to FAD with the production of 1.5 molecules of ATP.

The trans-alkene undergoes hydration to form a secondary alcohol in the second step. The catalyst is the enzyme enoyl-CoA hydratase — a stereospecific enzyme yielding only the L isomer. Next, the secondary alcohol undergoes oxidation to form a ketone. The oxidizing agent is NAD^+. The enzyme catalyzing this oxidation is β-hydroxy-acyl-CoA dehydrogenase. The re-oxidation of NADH to NAD^+ via the electron transport chain produces two molecules of ATP.

The final step involves the cleavage of the β-ketoacyl-CoA with a molecule of CoA. This produces acetyl-CoA and a fatty acyl-CoA two carbon atoms shorter than the original. The enzyme from this step is β-ketothiolase (or simply thiolase). The new fatty acyl-CoA goes around the cycle to be shortened by two carbon atoms. An unsaturated fatty acid also goes through similar steps, but needs one or two additional enzymes.

The energy yield from a fatty acid is larger than from glucose. The process begins with the activation of the fatty acid, which costs the equivalent of two ATP molecules. Each trip around the cycle yields ten molecules of ATP, a molecule of $FADH_2$, and a molecule of NADH. The NADH and $FADH_2$ ultimately yield four additional molecules of ATP. Thus, each trip around the cycle produces 14 molecules of ATP. In addition, the final trip around the cycle produces not one but two molecules of acetyl-CoA.

Not so heavenly bodies: Ketone bodies

Some of the excess acetyl-CoA will form a group of relatively small molecules called *ketone bodies.* This is especially important when there is a build up of acetyl-CoA. A build up may occur when the rate of production is too high or if it is not used efficiently. Two acetyl-CoA molecules combine in the reverse of the last step in β-oxidation to produce acetoacetyl-CoA. Acetoacetyl-CoA reacts with water and another acetyl-CoA to form β-hydroxy-β-methylglutaryl-CoA, which in turn cleaves to acetoacetate and acetyl-CoA. Most of the acetoacetate undergoes reduction to β-hydroxybutyrate (a small amount decarboxylates to acetone and carbon dioxide). These steps appear in Figure 13-17.

The other guy

When a fat molecule breaks down, the results are a glycerol and three fatty acid molecules. The fatty acid molecules enter the β-oxidation cycle and produce energy. Catabolism of the glycerol also serves as a source of energy. First, the glycerol is phosphorylated to glycerol 1-phosphate (= glycerol 3-phosphate). This uses one molecule of ATP. Oxidation of glycerol 1-phosphate generates dihydroxyacetone phosphate, which can enter the glycolysis pathway. The net energy production is from 16.5 to 18.5 molecules of ATP.

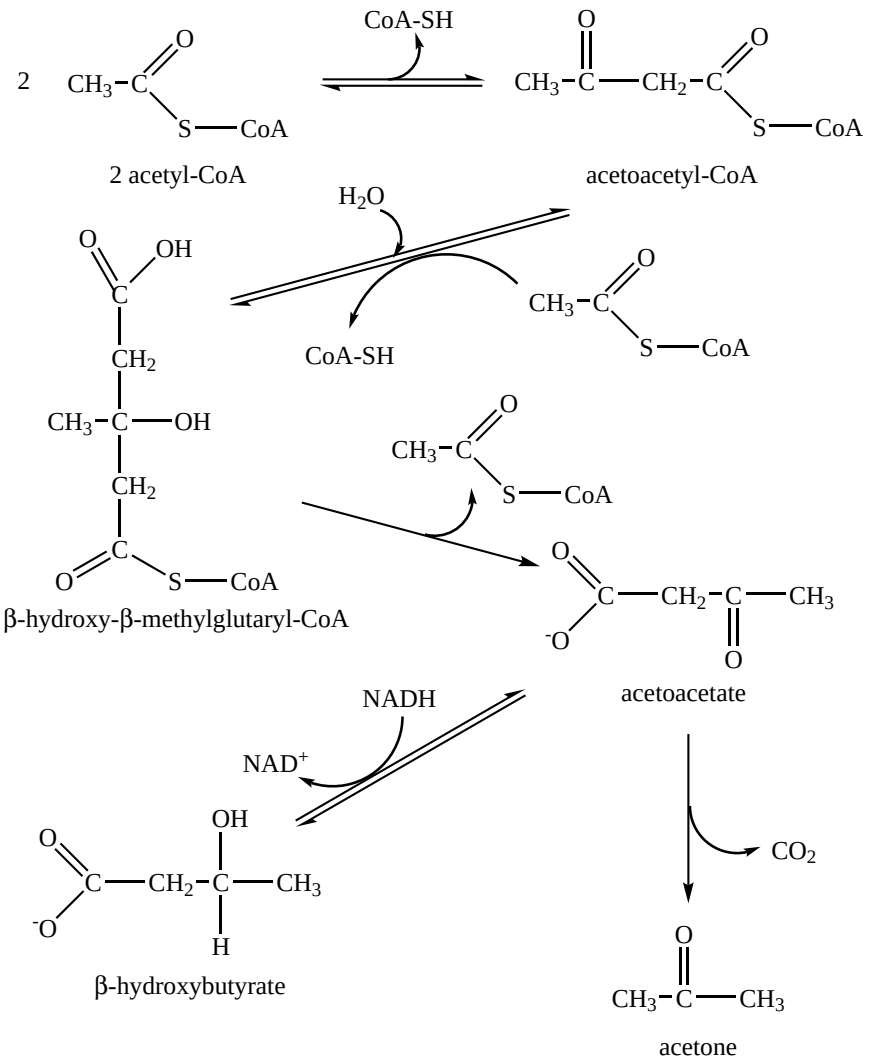


Figure 13-17: Formation of the ketone bodies.

As a group, acetone, β-hydroxybutyrate, and acetoacetate are the ketone bodies.

This process occurs primarily in the liver, and the β-hydroxybutyrate and acetoacetate then enter the bloodstream for use by other tissues. During prolonged starvation, ketone bodies may serve as the major energy source for some tissues. The kidneys excrete excess ketone bodies. Normal blood levels are about 1 mg of ketone bodies per 100 mL of blood.

In starvation or diabetes mellitus, a form of diabetes, cells may not receive sufficient carbohydrate for energy, leading to an increase in the rate of fatty acid oxidation to compensate for the energy deficit. As the amount of acetyl-CoA increases, there is insufficient oxaloacetate in the citric acid cycle available for oxidation of this acetyl-CoA. (The oxaloacetate concentration is lower because of the necessity of using it for glucose synthesis.) This leads to an increase in the production of ketone bodies and an increase of ketone bodies in the bloodstream. At 3 mg of ketone bodies per 100 mL, a condition known as *ketonemia* arises — a high concentration of ketone bodies in the urine. Ketonemia and ketonuria are two aspects of ketosis.

Two of the ketone bodies are in the form of acids. The build up of ketone bodies leads to an overwhelming of the blood buffers. The decrease in blood pH may reach 0.5 units lower than the normal pH (7.4), leading to acidosis, a serious condition, which, among other things, leads to difficulty in oxygen transport by hemoglobin. Dehydration results as the kidneys eliminate large quantities of liquid trying to remove the excess acid. Severe acidosis may result in a coma that may result in death.

Mammals cannot convert acetyl-CoA to carbohydrates. It is possible to convert carbohydrates to fats, but not to do the reverse.

Investing in the Future: Biosynthesis

One aspect of metabolism, catabolism, is to produce the energy required for life. Another aspect, anabolism, is to supply the materials for growth and replacement. Food supplies the raw fuel for metabolism. A number of pathways are available to allow for flexibility. It is necessary to block some pathways to overcome Le Châtelier's Principle, partly because an enzyme will catalyze both the forward and the reverse reaction.

Nearly all intermediates in catabolic processes are also intermediates in anabolic processes. Molecules may easily change from one pathway to another. In general, anabolic processes require the energy produced by catabolic processes. We've already seen one aspect of anabolism — gluconeogenesis. Earlier, we saw how this process, related to glycolysis, could generate glucose and other carbohydrates. We examine other biosynthesis processes in this section.

Fatty acids

Production of the fatty acids is necessary to form the membrane lipids. But the main reason for fatty acid synthesis is to convert excess dietary carbohydrate to fats for storage. The key molecule for this is acetyl-CoA.

The liver is the primary fatty acid synthesis site in humans, and humans can synthesize all the fatty acids but two: linoleic acid and linolenic acid. Linoleic acid and linolenic acid are also essential fatty acids, required components of the diet. Acetyl-CoA from glycolysis or β-oxidation reacts with bicarbonate ion in a reaction (Figure 13-18) powered by ATP and catalyzed by acetyl-CoA carboxylase, forming the three-carbon molecule malonyl-CoA.

The release of insulin triggers a series of steps that result in the activation of acetyl-CoA carboxylase. Release of insulin indicates high food levels. Both glucagon and epinephrine inhibit the enzyme, through a series of steps. In mammals, the enzymes necessary to synthesize palmitic acid from acetyl-CoA and malonyl-CoA are present in a complex known as *fatty acid synthase.* In plants and bacteria, the enzymes are present as separate molecules. Synthesis proceeds two carbon atoms at a time, which is why all the natural fatty acids contain an even number of carbon atoms.

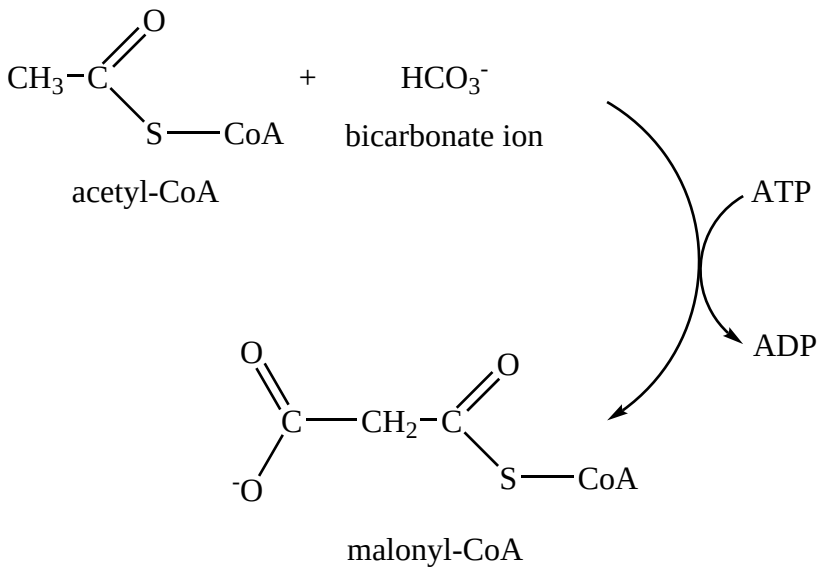


Figure 13-18: Synthesis of malonyl-CoA.

Synthesis begins when a molecule of acetyl-CoA links to an acyl carrier protein, ACP, and a malonyl-CoA does the same with another ACP. The two ACP-linked molecules then join and release a carbon dioxide molecule, an ACP, and an acetoacetyl-ACP. Next are three steps that are the reverse of the first three steps of β-oxidation. First, NADPH reduces the ketone group to an

alcohol. Then dehydration of the alcohol leaves a double bond between the second and the third carbon atoms. The coenzyme NADPH again serves as a reducing agent to produce butyryl-ACP. The sequence repeats with butyryl-ACP replacing the acetyl-ACP. These steps are in Figure 13-19.

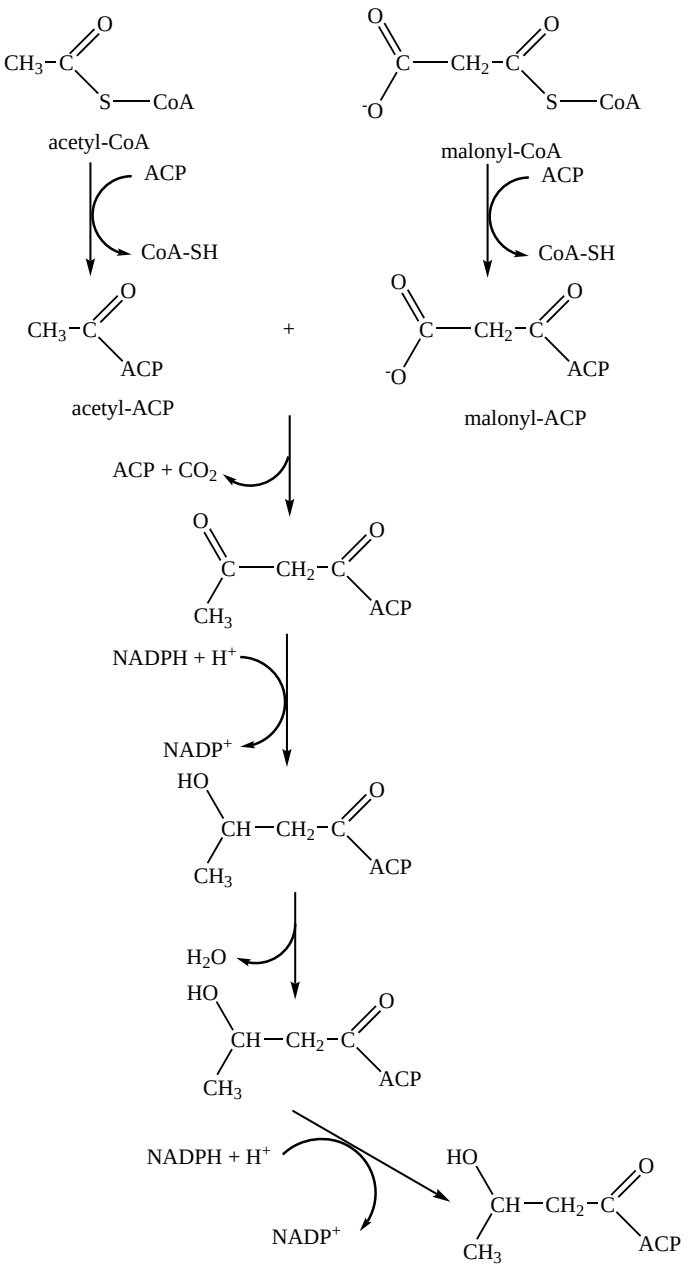


Figure 13-19: Fatty acid synthesis.

The series of synthesis steps continues up to palmitic acid (16 carbon atoms). The overall reaction is:

$$8 \text{ acetyl-CoA} + 7 H^+ + 14 \text{ NADPH} + 7 \text{ ATP} \rightarrow \text{palmitic acid} + 8 \text{ CoA} + 14 \text{ NADP}^+ + 7 \text{ ATP} + 7 P_i$$

Once the palmitic acid forms, additional reactions, where necessary, can lengthen or shorten the chain. These require different enzyme systems. Partial oxidation of a saturated fatty acid yields an unsaturated fatty acid.

Membrane lipids

Like other molecules, it is necessary to synthesize the membrane lipids from their constituents. In the previous section, we explained how to synthesize the fatty acids. These fatty acids need to be activated with acetyl-CoA in order to produce the appropriate acyl-CoA. The reduction of dihydroxyacetone, from glycolysis, yields glycerol 3-phosphate. The glycerol 3-phosphate combines with the appropriate acyl-CoA molecules to yield a phosphatidate (Figure 13-20). The phosphitidate then reacts with an activated serine or an activated choline to form the appropriate phosphoglyceride.

The formation of the spingolipids follows a similar path. In this case, sphingosine replaces glycerol. The synthesis of sphingosine begins with the reaction of palmitoyl-CoA, with serine in the presence of acid. This reaction yields Coenzyme A, carbon dioxide, and the precursor of sphingosine. Oxidation of the precursor yields sphingosine (Figure 13-21).

An acyl-CoA can then add a fatty acid to the amine group to produce N-acylsphingosine (ceramide). The reaction of the alcohol on the third carbon of the ceramide with activated phosphocholine yields sphingomyelin.

The reaction of ceramide with an activated monosaccharide begins the synthesis of the glycolipids. To complete the synthesis, it is necessary to add additional activated monocaccharides (UDP-glucose being one example).

Cholesterol is another membrane lipid. It helps to control the fluidity of cell membranes and is also the precursor of the steroid hormones. The entire synthesis takes place in the liver, where acetyl-CoA molecules are joined. Thus, the cholesterol molecule is built up two carbon atoms at a time.

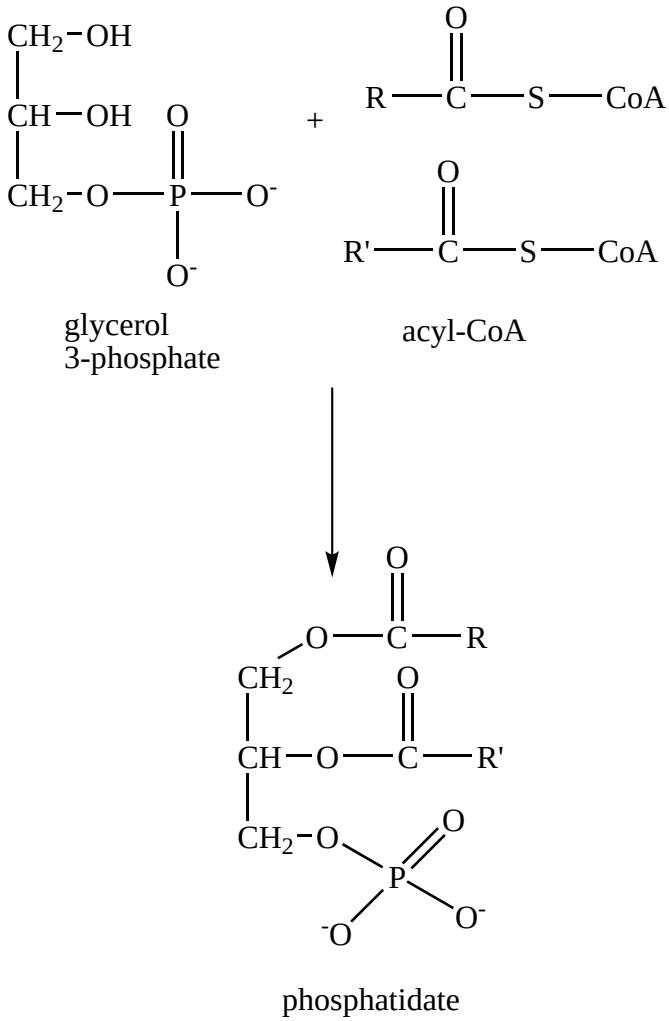


Figure 13-20: Formation of phosphatidate.

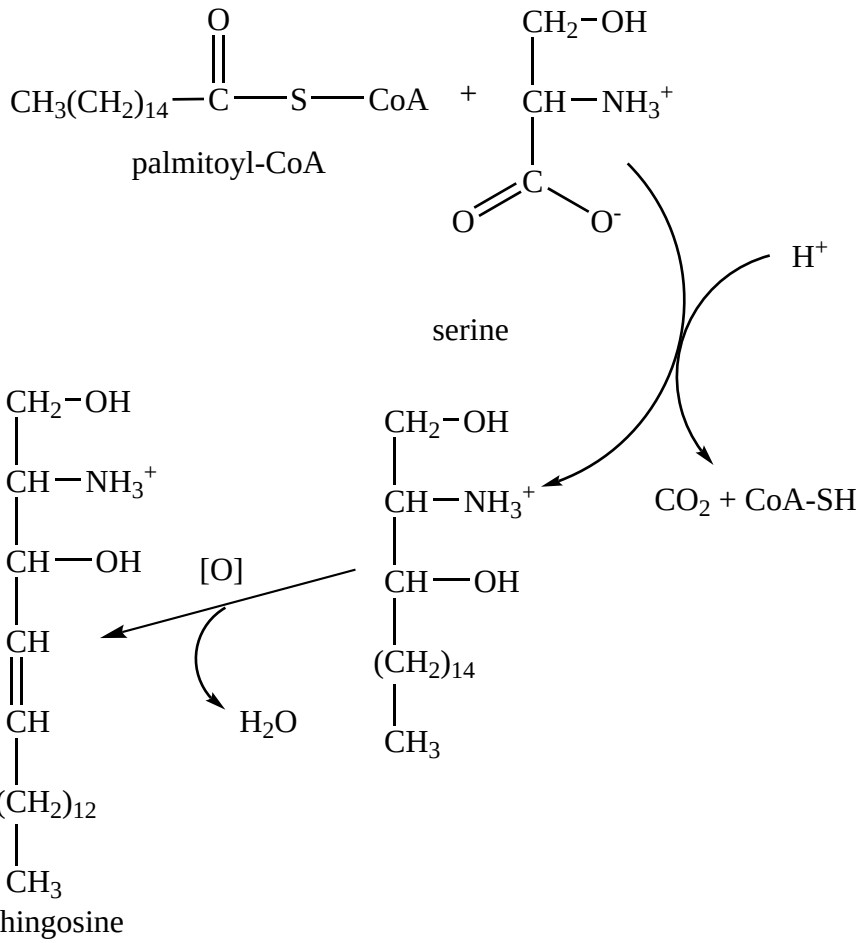


Figure 13-21: Formation of sphingosine.

Amino acids

Synthesis of amino acids becomes necessary when insufficient quantities are present in the diet. Adult humans can only synthesize 11 of the 20 amino acids. The amino acids that humans cannot synthesize are known as the *essential amino acids,* and these are a necessary requirement in the diet. Table 13-2 list the essential and non-essential amino acids.

Table 13-2 Essential and Non-essential Amino Acids

Essential	*Non-essential*
Histidine	Alanine
Isoleucine	Asparagine
Leucine	Aspartate
Lysine	Cysteine
Methionine	Glutamine
Phenylalanine	Glutamate
Threonine	Glycine
Tryptophan	Proline
Valine	Serine

Arginine is essential for children, but not for adults. Tyrosine is non-essential in the presence of adequate quantities of phenylalanine. Glutamate is important to the synthesis of five amino acids. Glutamate may form by the reduction of α-ketoglutaric acid, an intermediate from the Krebs cycle. The process is shown in Figure 13-22.

In the forward direction, this is a synthesis reaction, whereas the reverse reaction is an important oxidative deamination from the catabolism of amino acids. Glutamate, when necessary, serves as an intermediate in the biosynthesis of alanine, aspartate, asparagine, glutamine, proline, and serine. The transamination in Figure 13-23 illustrates the formation of alanine.

Replacing pyruvate in the preceding reaction with oxaloacetate yields aspartate.

It is possible to convert excess phenylalanine to tyrosine by a simple oxidation in the presence of phenylalanine hydroxylase (Figure 13-24).

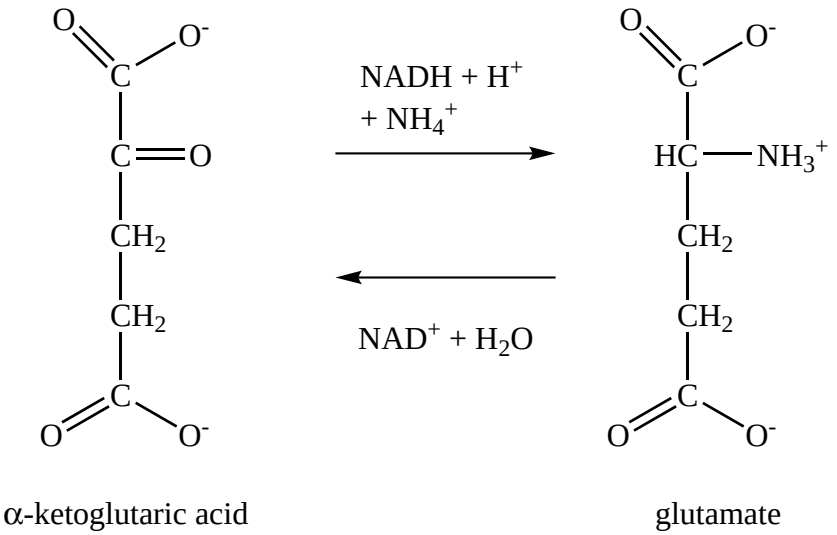


Figure 13-22: Equilibrium between glutamate and α-ketoglutaric acid.

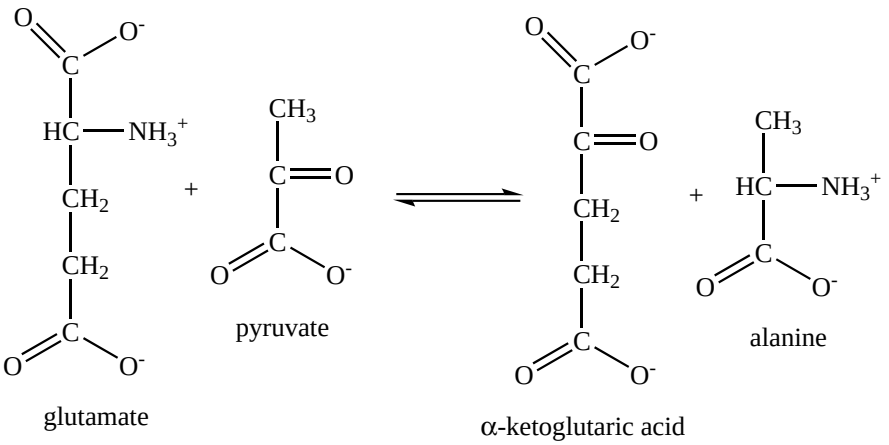


Figure 13-23: Formation of alamine.

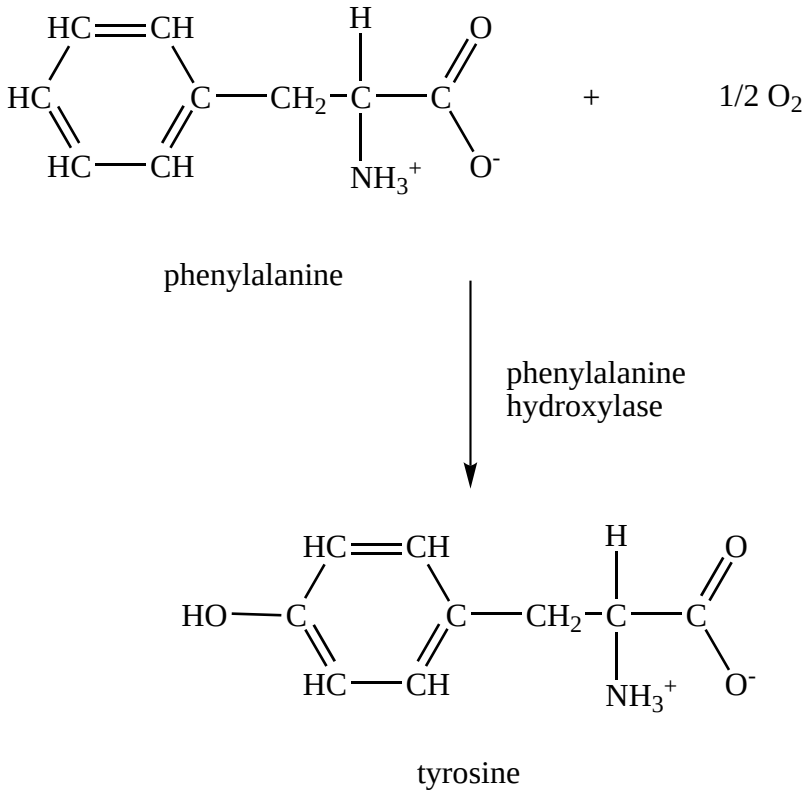


Figure 13-24: Synthesis of tyrosine.

Methionine serves as the source of sulfur for the synthesis of cysteine. Serine serves as the base of the rest of the molecule. *Serine* is the product of a three-step process beginning with 3-phosphoglycerate. The process starts with the oxidation by NAD^+ of the secondary alcohol group. The ketone thus formed undergoes transamination with glutamate to form 3-phosphoserine. Finally, hydrolysis of the phosphate ester yields serine (Figure 13-25).

The formation of proline is a four-step process beginning with glutamate. The process is shown in Figure 13-26.

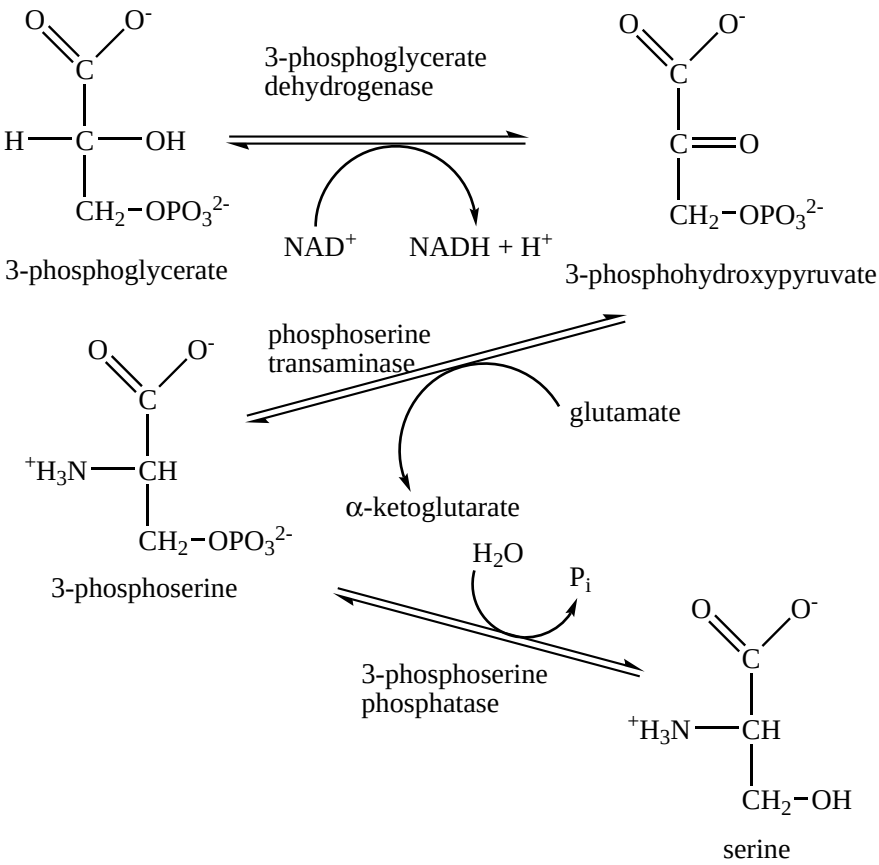


Figure 13-25: Synthesis of serine.

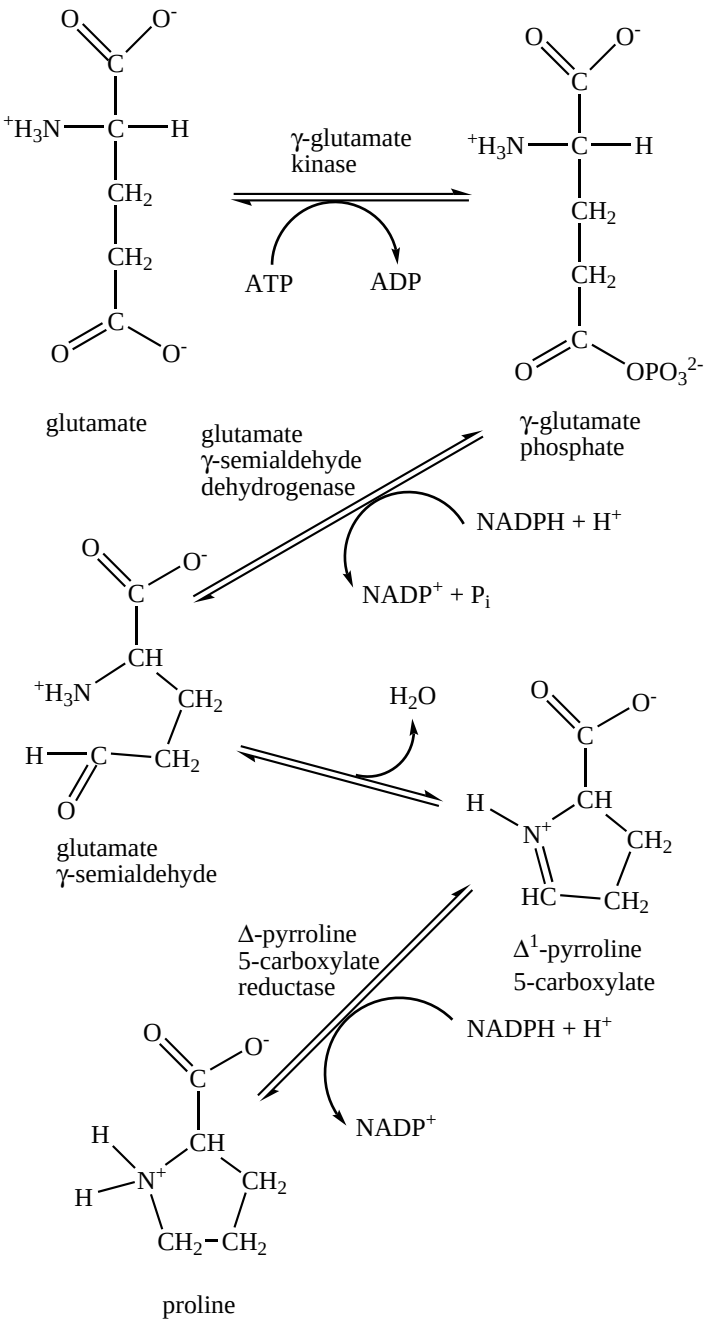


Figure 13-26: Synthesis of proline.

Chapter 14

Smelly Biochemistry: Nitrogen in Biological Systems

In This Chapter

- Talking about purine and pyrimidine
- Examining catabolism and discussing the urea cycle
- Considering amino acids
- Finding out about metabolic disorders

In this chapter, we investigate the role of nitrogen in biomolecules. Nitrogen occurs primarily in the amino acids (proteins) and in nucleic acids (purines and pyrimidines), many of which have a distinctive and generally unpleasant aroma, hence our chapter title. A few other molecules, such as hemoglobin, also contain nitrogen. Humans eliminate nitrogen primarily in the urea.

Ring in the Nitrogen: Purine

Adenine and guanine are nitrogen bases that employ the purine ring system (Figure 14-1). The formation of these molecules is essential to the synthesis of both DNA and RNA. The biosynthesis of the purines generates the molecules in their nucleotide forms instead of the free base form.

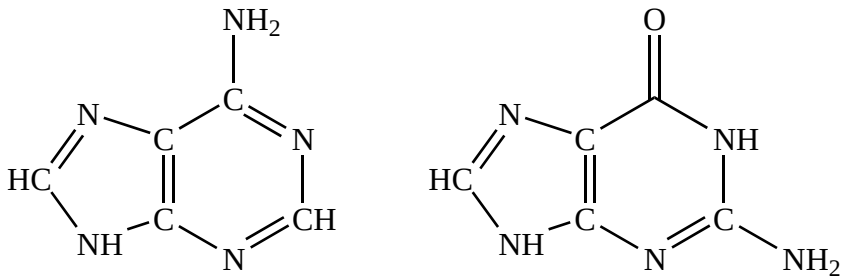


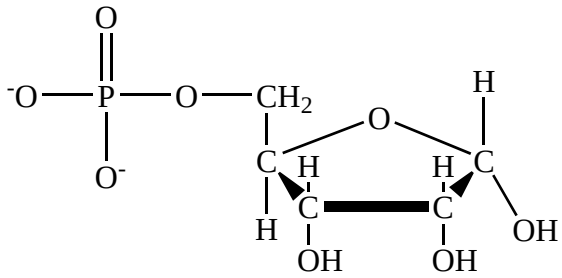
Figure 14-1: Purine nitrogen bases.

Biosynthesis of purine

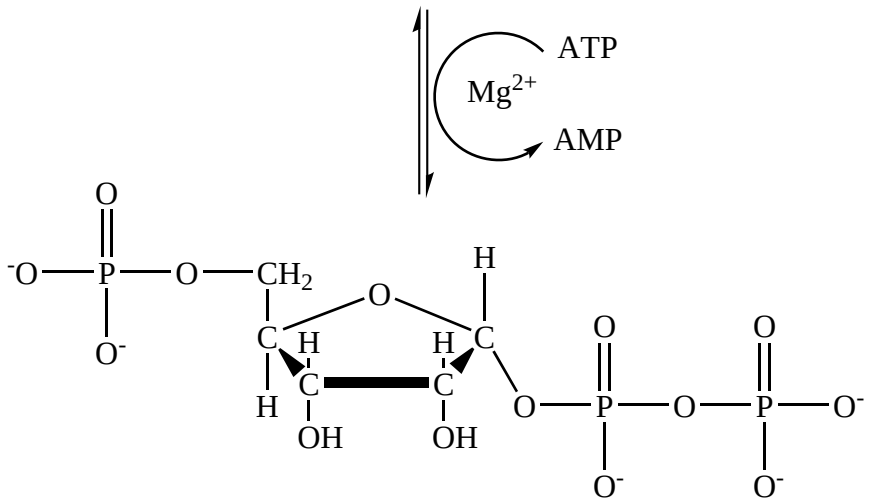
The synthesis of purine begins with the activation of D-ribose-5'- phosphate through pyrophosphorylation. In this reaction a pyrophosphate group from ATP is transferred to C-1 of an α-D-ribose-5'-phosphate. This gives a 5-phospho-α-D-ribose 1-pyrophosphate (PRPP) and AMP. The reaction is unusual because it involves the transfer of an intact pyrophosphate group (Figure 14-2). PRPP is also necessary for the synthesis of pyrimidines.

Inosine synthesis

PRPP goes through a series of ten steps (Figure 14-3) to become inosine 5'-phosphate or inosinic acid (IMP). Notice that throughout these ten steps the D-ribose-5'-phosphate portion of PRPP does not change. The ten enzymes necessary for these steps are in Table 14-1. Two additional, though different, steps are necessary to convert IMP to either AMP or GMP.



α-D-ribose 5-phosphate



5-phospho-α-D-ribose 1-pyrophosphate (PRPP)

Figure 14-2: Activation of D-ribose-5'-phosphate.

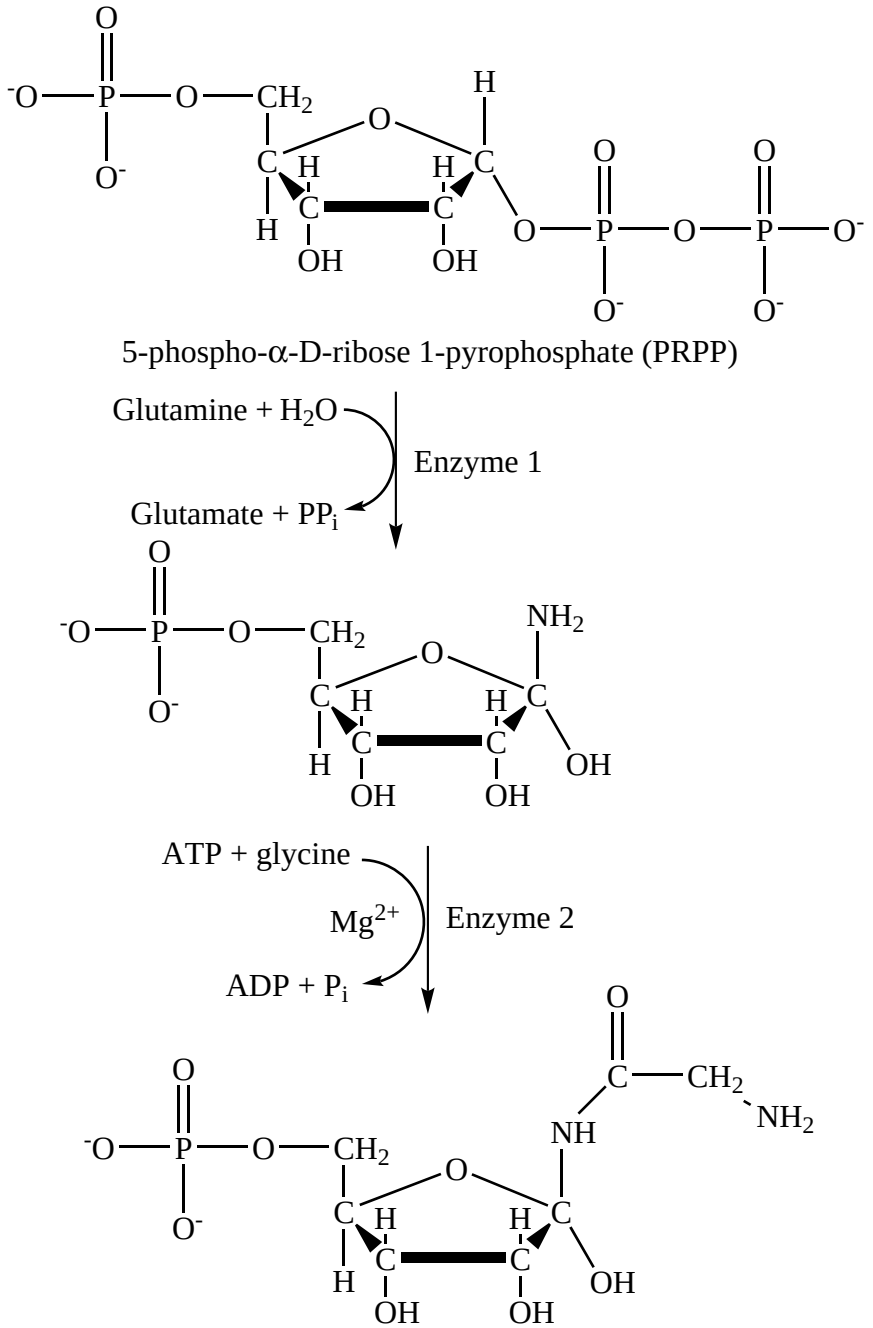
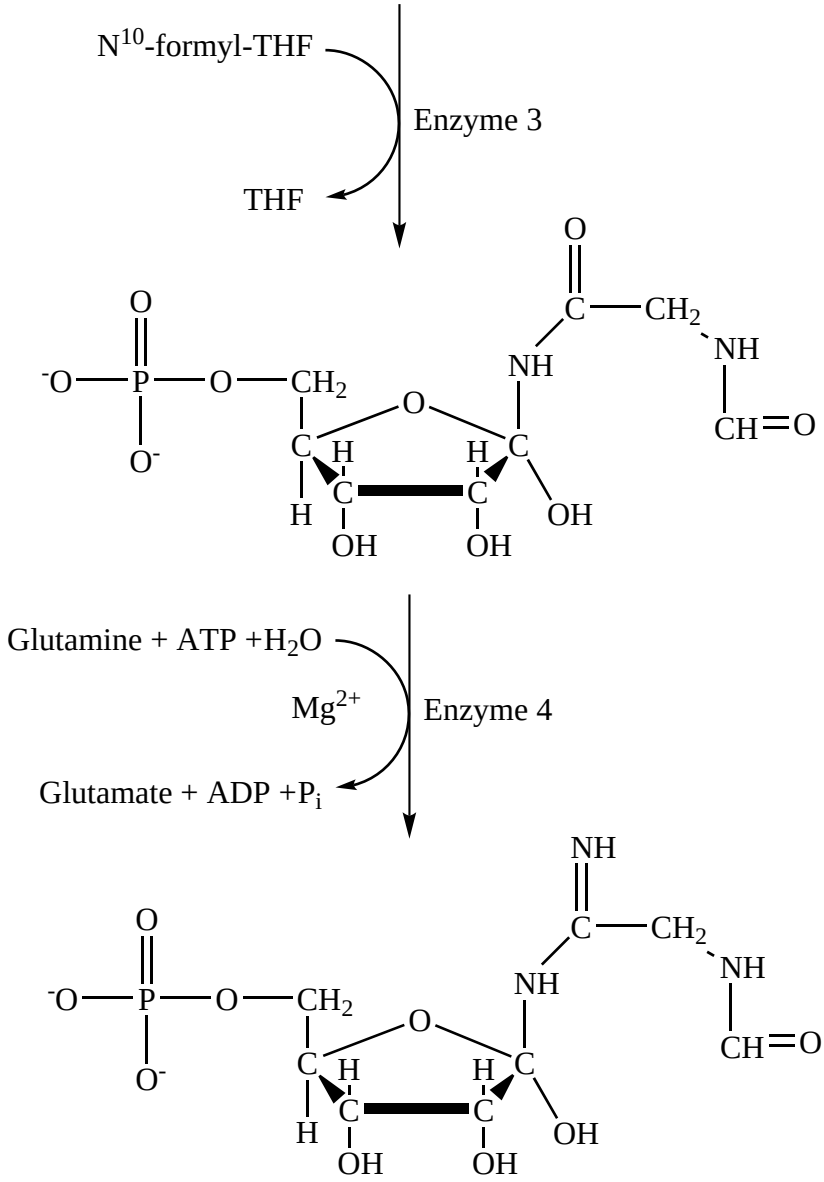
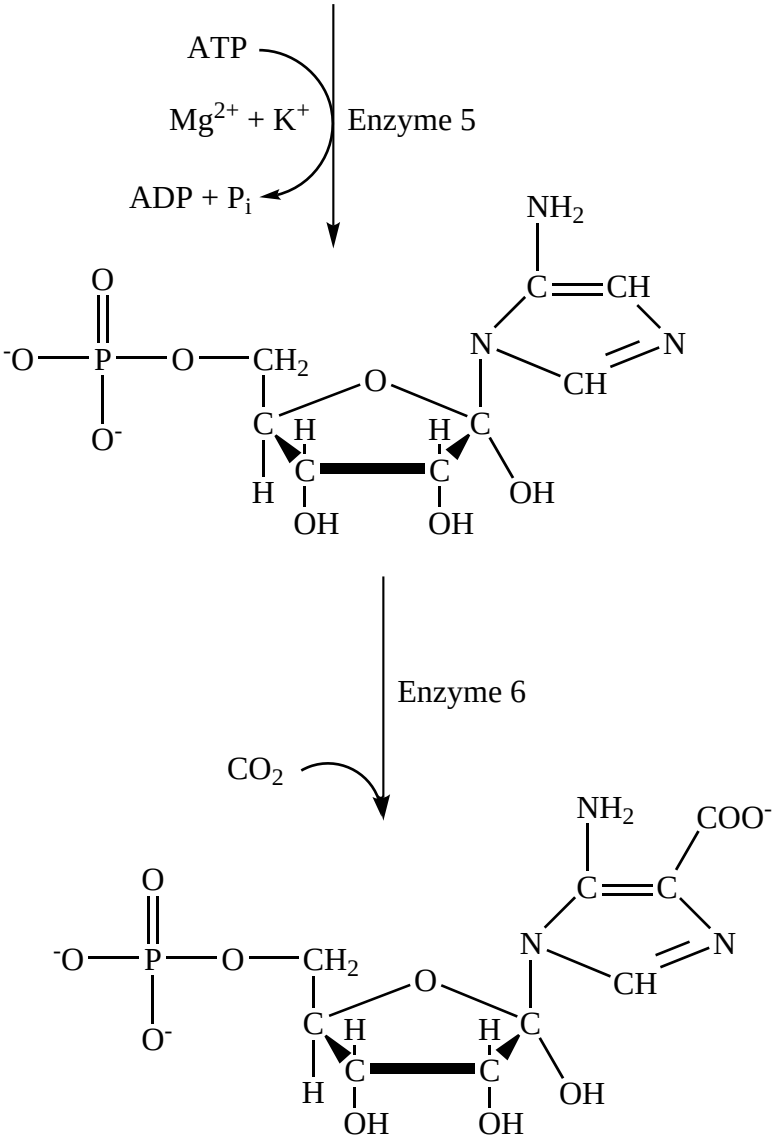


Figure 14-3: The ten steps necessary to convert PRPP (5-phospho-α-D-ribose 1-pyrophosphate) into inosine 5'-phosphate.


N10-formyl-THF
Enzyme 3
THF
Glutamine + ATP +H2O
Mg2+
Enzyme 4
Glutamate + ADP +Pi


ATP
Mg2+ + K+
Enzyme 5
ADP + Pi
NH2
O
-O
P
O
CH2
O
O-
C
CH
N
N
CH
C
H
H
C
H
C
C
H
OH
OH
OH
Enzyme 6
CO2
NH2
COO-
C
C
N
N
CH

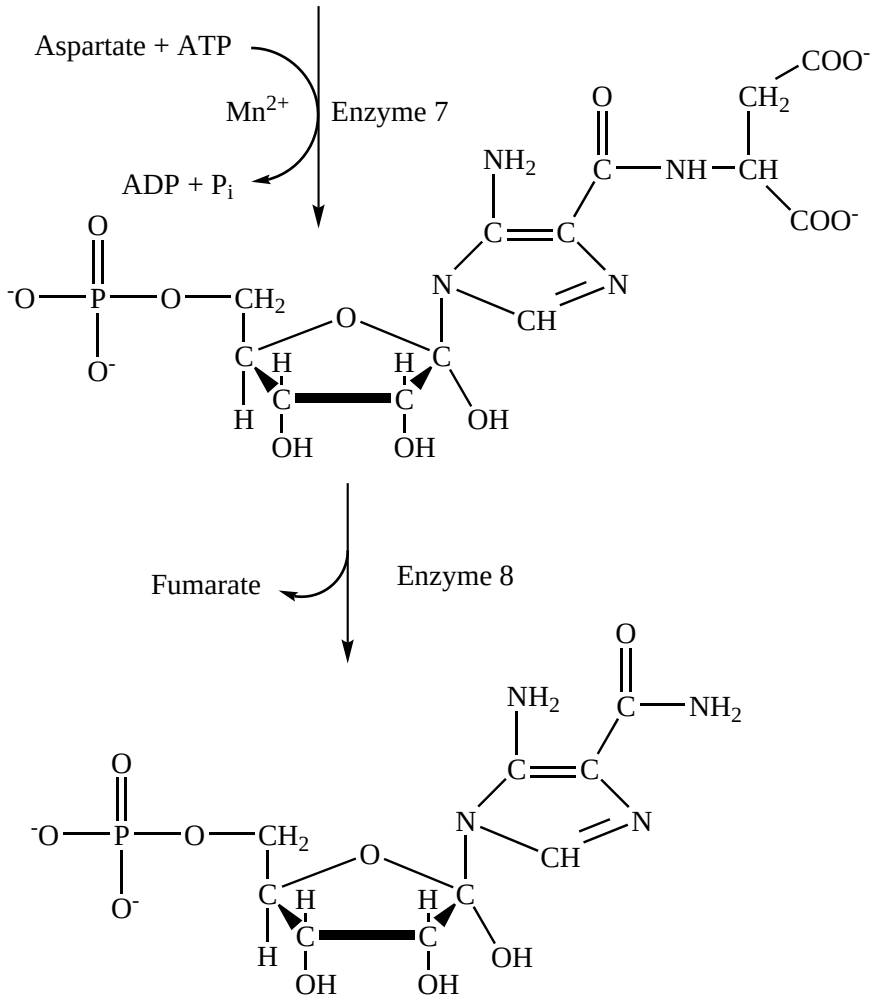

Aspartate + ATP
Mn2+
Enzyme 7
ADP + Pi
Fumarate
Enzyme 8

Table 14-1 Ten Enzymes Necessary for Inosine Synthesis

Enzyme	*Name*
1	Amidophosphoribosyl transferase
2	Phosphoribosylglycinamide synthetase
3	Phosphoribosylglycinamide formyltransferase
4	Phosphoribosylformylglycimamide synthetase
5	Phosphoribosylaminoimidazole synthetase
6	Phosphoribosylaminoimidazole carboxylase
7	Phosphoribosylaminoimidazole-succinocarboxamide synthesase
8	Adenylosuccinate lyase
9	Phosphoribosylaminoimadazolecarboxamide formyltransferase
10	IMP cyclohydrolase

AMP synthesis

To convert IMP into AMP, it is necessary to transfer an amino group from an aspartate. This transfer requires two steps, and the energy to add aspartate to IMP comes from the hydrolysis of a GTP. The process is then completed by the loss of fumarate. The enzyme adenylosuccinate synthetase catalyzes the first step, and the enzyme adenylosuccinate lyase catalyzes the second. Figure 14-4 illustrates the process.

GMP synthesis

The conversion of IMP to GMP begins with the IMP dehydrogenase catalyzed oxidation to xanthosine 5'-phosphate. The coenzyme for this step is NAD^+. GMP synthetase catalyzes the next step — the amine transfer from glutamate. The energy for this step is supplied by the hydrolysis of ATP (Figure 14-5).

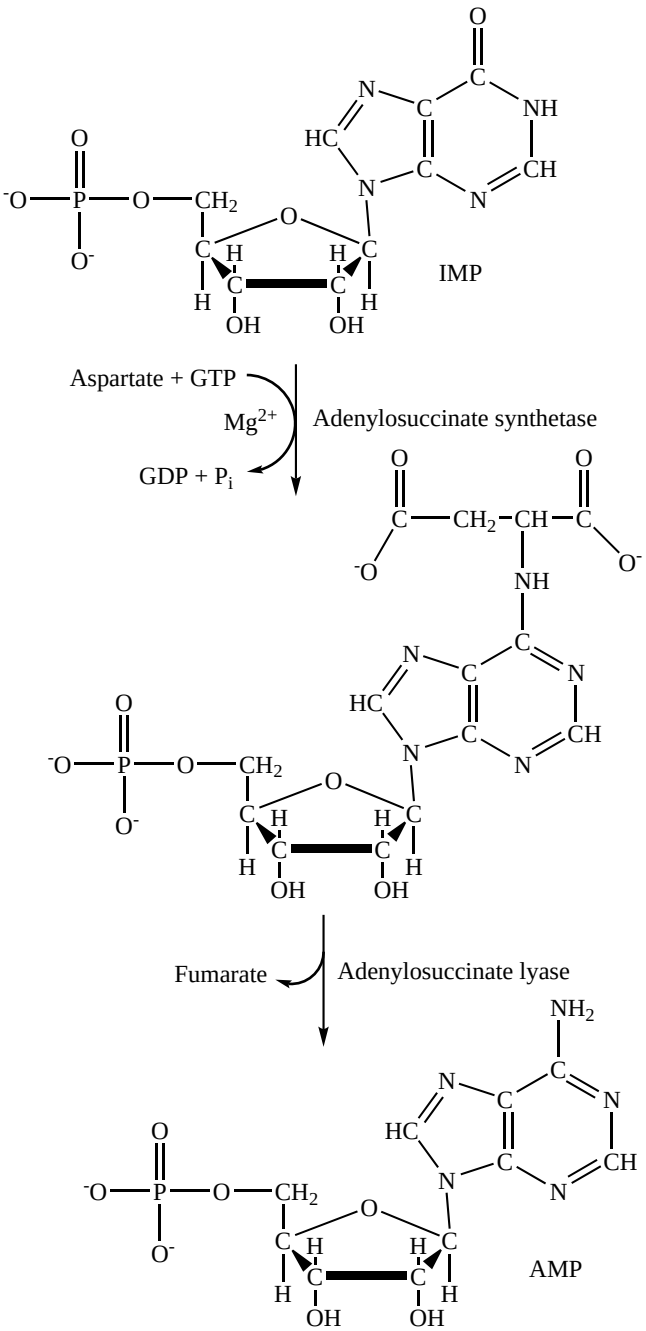


Figure 14-4: Conversion of IMP to AMP.

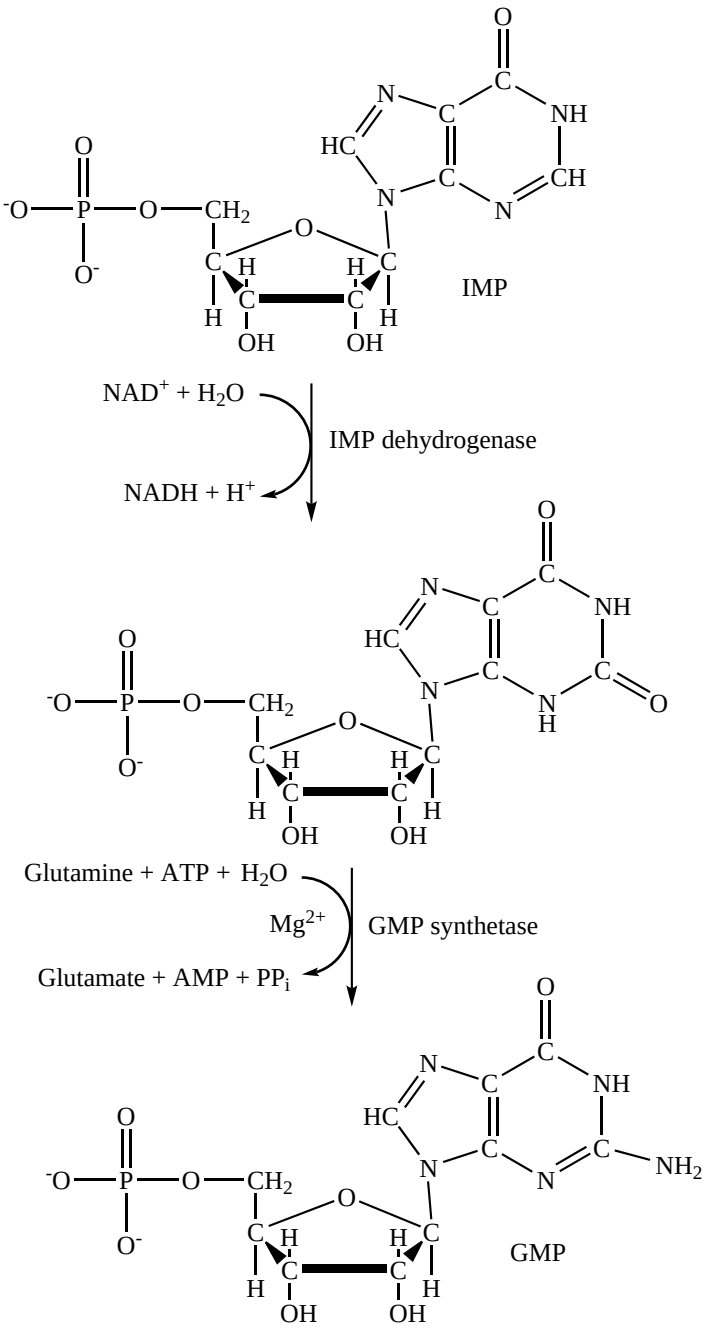


Figure 14-5: Conversion of IMP to GMP.

How much will it cost?

The biosynthesis of both AMP and GMP requires the hydrolysis of several high-energy bonds. To produce IMP from D-ribose 5-phosphate requires the hydrolysis of five high-energy bonds (one PP_i and five ATP). To convert IMP to AMP requires the hydrolysis of one more high-energy bond (from GTP). And to convert IMP to GMP requires the hydrolysis of two high-energy bonds — one ATP and one PP_i.

Anaerobic organisms, such as the bacteria responsible for tetanus or botulism, must oxidize four glucose molecules at two ATP per glucose to meet the energy requirement. An aerobic organism, like you, for example, needs to oxidize only one glucose molecule at 36 or 38 ATP per glucose. The preceding processes require a substantial amount of energy. Sometimes this energy requirement may be lessened by metabolic processes known as the *salvage pathways.* In the salvage pathways, nitrogen bases are recycled instead of synthesized. The nitrogen bases are then converted to nucleotides.

Pyrimidine Synthesis

The biosynthesis of pyrimidines follows a different path from purine synthesis. In this case, synthesis of the base takes place before attachment to the ribose. Ring synthesis requires bicarbonate ion, aspartic acid, and ammonia. Although it is possible to use ammonia directly, it usually comes from the hydrolysis of the side chain of glutamine.

First step: Carbamoyl phosphate

The initial step is to transfer a phosphate from an ATP to a bicarbonate ion to form carboxyphosphate, which in turn undergoes an exchange where ammonia replaces the phosphate to form carbamic acid. Whew! — got that? A second ATP transfers a phosphate to carbamic acid to form carbamoyl phosphate. Figure 14-6 summarizes these steps.

The primary enzyme for the process in Figure 14-6 is carbamoyl synthetase. One region of the enzyme is responsible for the synthesis of carbamic acid, whereas a second region hydrolyzes ammonia from glutamine. A third region completes the process, and a channel connects the three regions.

Next step: Orotate

The next step in pyrimidine synthesis is the formation of orotate, which will be joined to a ribose. It begins with the enzyme aspartate transcarbamoylate,

which joins aspartate to carbamoyl phosphate with the loss of phosphate. This forms carbamoylaspartate. Carbamoylaspartate cyclizes to dihydroorotate, which is oxidized by NAD^+ to orotate (Figure 14-2).

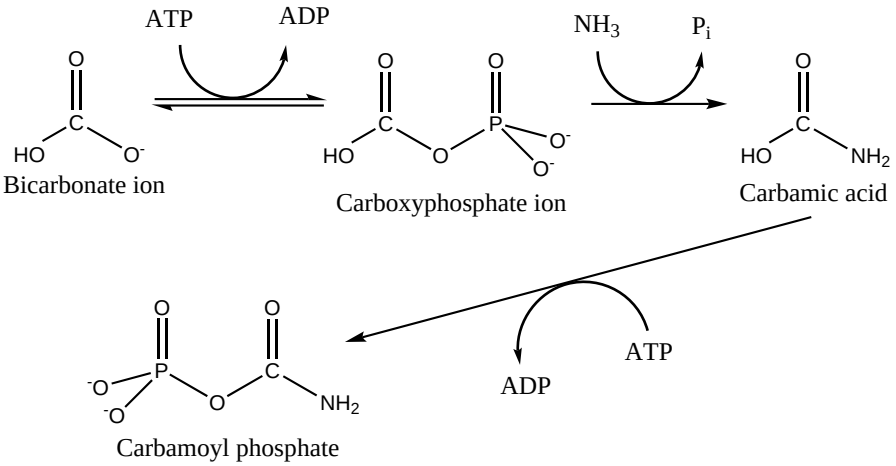


Figure 14-6: Synthesis of carbamoyl phosphate.

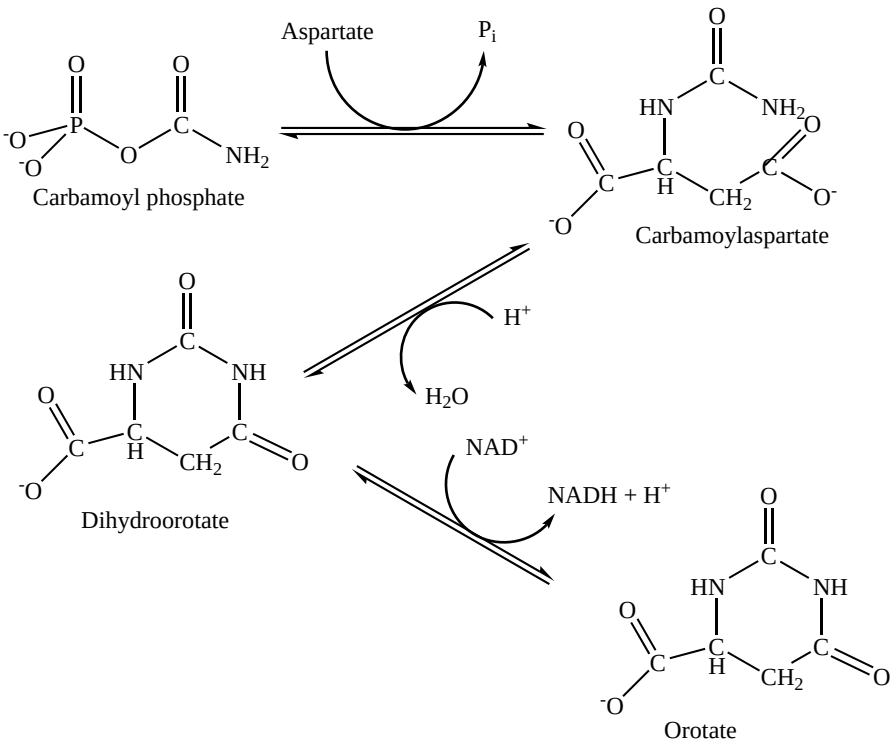


Figure 14-7: Formation of orotate from carbamoyl phosphate.

Orotate joins with 5-phosphoribosyl-1-pyrophosphate (PRPP) to form orotidylate, with pyrophosphate hydrolysis providing the energy necessary. The enzyme pyrimidine phosphoribosyltransferase is responsible for this reaction. The enzyme orotidylate decarboxylase catalyzes the decarboxylation of orotidylate to uridylate (UMP). Figure 14-8 illustrates these steps.

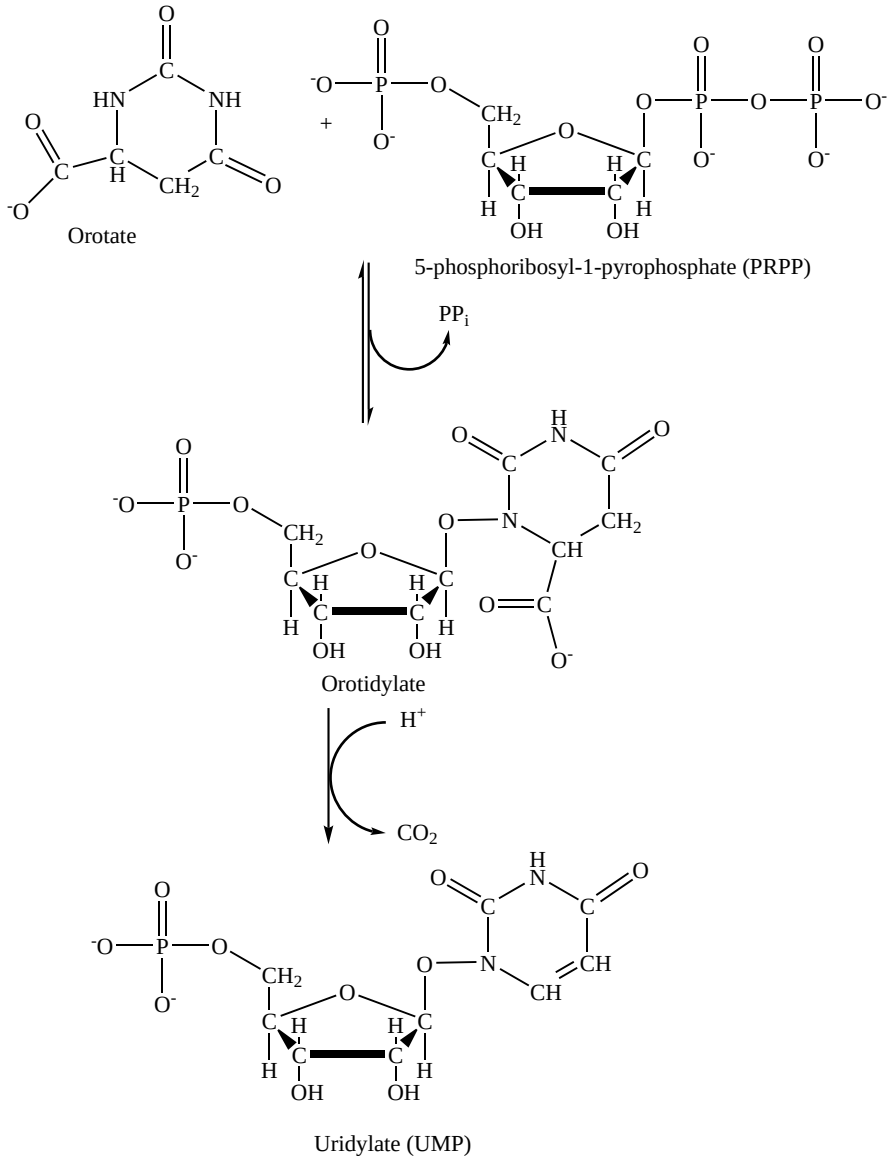


Figure 14-8: Conversion of orotate to uridylate (UMP).

Last step: Cytidine

The final nucleotide, cytidine, forms from uridinemonophosphate (UMP). The first step is to change UMP into UTP. UMP kinase transfers a pyrophosphate from ATP to UMP. Figure 14-9 shows this process.

Back to the Beginning: Catabolism

Catabolism, remember, is the breaking down of molecules to provide energy. In many cases, a complete breakdown is not necessary, because the products from a partial breakdown can be reused when necessary.

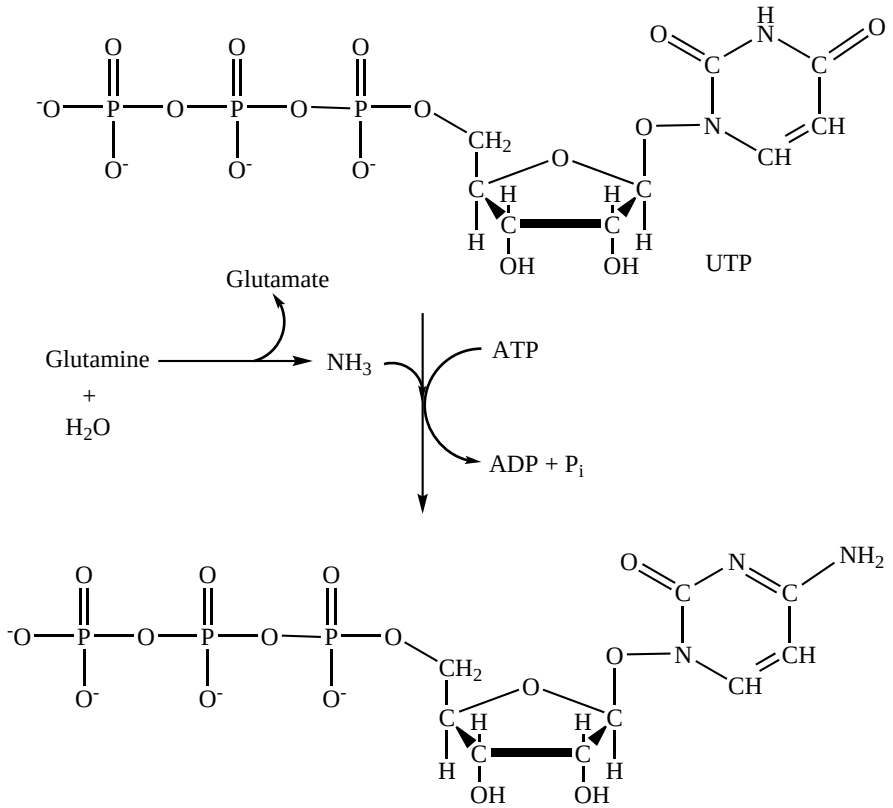


Figure 14-9: Conversion of UTP to CTP.

Nucleotide catabolism

The breakdown of the nucleotides begins with the removal of a phosphate group (from C-5). Next, a phosphate attaches to C-1 to give the sugar-1-phosphate, and the base leaves. In humans and many other species, uric acid (Figure 14-10) is the product of further degradation of purines. Other biochemical species further degrade uric acid into other products.

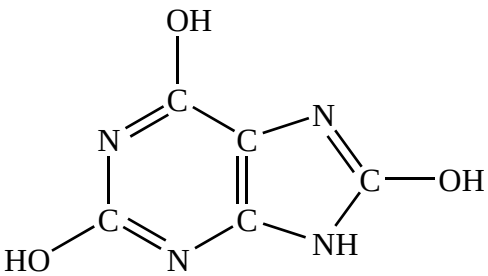


Figure 14-10: Structure of uric acid.

Amino acid catabolism

Hydrolysis of proteins yields the separate amino acids. It is possible to recycle these amino acids, use them in the synthesis of other amino acids, or produce energy from them. Through transamination it is possible to transfer an amino group from any amino acid (other than lysine, proline, or threonine), and an α-keto acid. The general category of enzymes that catalyzes this reaction is a transaminase, and the general reaction is shown in Figure 14-11. Nitrogen destined for elimination transfers to α-ketoglutarate to form glutamate. Transamination is important in the biosynthesis of alanine, aspartate, and glutamate.

Oxidative deamination of glutamate forms α-ketoglutarate (to be recycled), an ammonium ion (to enter the urea cycle) and, indirectly, 3 ATP. Glutamate dehydrogenase and either NAD^+ or $NADP^+$ are necessary for this.

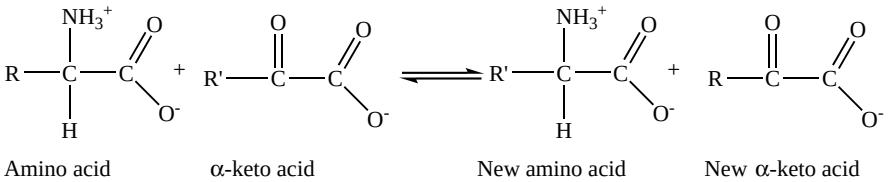


Figure 14-11: General transamination reaction.

The deaminated amino acid (α-keto acid) is further broken down. The α-keto acid may be broken down to pyruvate or some other material the body can

use to form glucose. These acids are called *glucogenic.* The alternative is to break down the α-keto acid to acetyl CoA and acetoacetic acid. These acids are called *ketogenic.* To further confuse you, some amino acids may be both glucogenic and ketogenic (see Table 14-2). These are the two possible fates of the carbon skeleton of the amino acids. The degradation of the amino acid transforms the carbon skeletons into intermediates in the citric acid cycle or into materials convertible to glucose.

Table 14-2 Glucogenic and Ketogenic Amino Acids

Glucogenic: Alanine, arginine, asparagine, aspartate, ccysteine, glutamate, glutamine, glycine, histidine, methionine, proline, serine, threonine, valine
Ketogenic: Leucine
Both: Isoleucine, lysine, phenylalanine, tyrosine, tryptophan

The general process is cyclic, with the various amino acids entering at different points. The basic scheme is shown in Figure 14-12.

Heme catabolism

The other important nitrogen compound in red-blooded organisms is *heme*. This species occurs in both hemoglobin and myoglobin. Hemoglobin is released as aged red blood cells are destroyed. The globin portion hydrolyzes to the appropriate amino acids. The iron separates from the heme and is stored in ferritin. Through a series of steps, bilirubin forms from the heme. The gall bladder temporarily stores bilirubin until the organism eliminates it.

Process of Elimination: The Urea Cycle

The catabolism of nitrogen-containing compounds yields recyclable nitrogen compounds and ammonia. Glutamine serves as temporary storage and transportation of the nitrogen — however, even small amounts of ammonia are toxic to humans. For this reason, ammonia must be converted to a less toxic form for elimination. The first step involves the conversion of ammonia, as the ammonium ion, to carbamoyl phosphate. The enzyme utilized for this conversion is carbamoyl phosphate synthetase. Figure 14-13 illustrates this reaction.

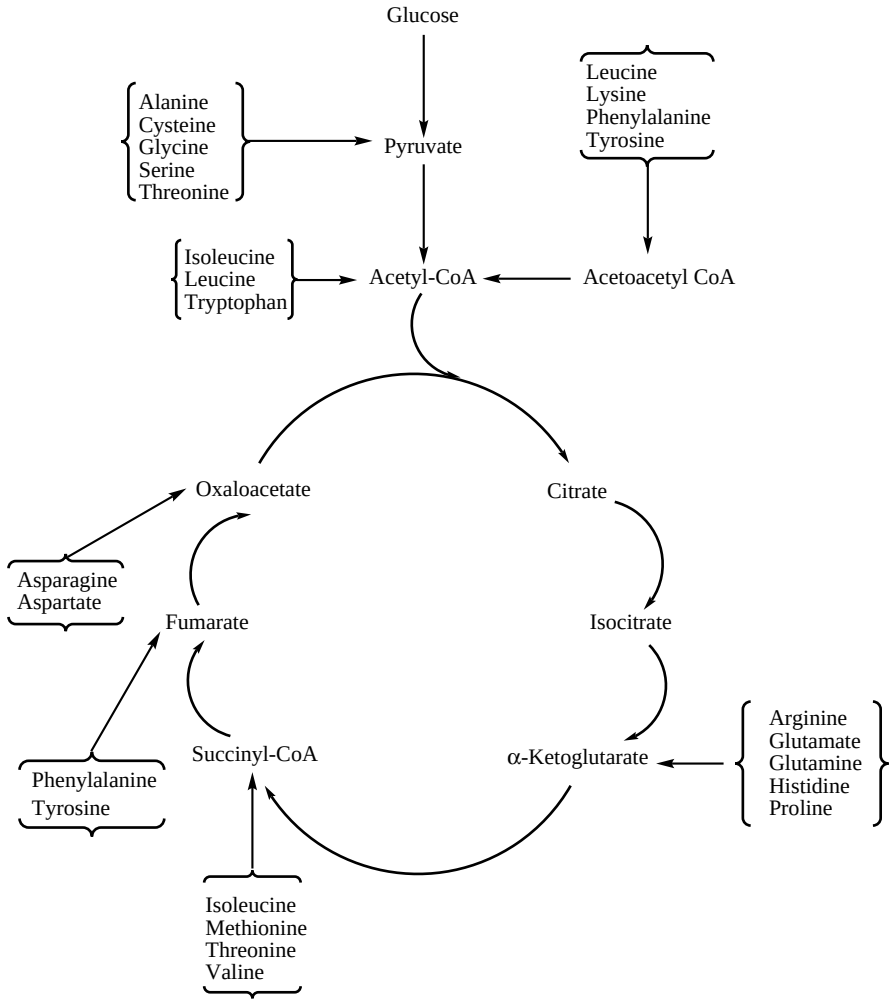


Figure 14-12: Fates of the amino acids.

Carbamoyl phosphate enters the urea cycle by joining to ornithine to produce citrulline, with the enzyme ornithine transcarbamoylase catalyzing this reaction. The enzyme arginosuccinate synthetase, with energy from the hydrolysis of ATP, joins aspartate to citrulline to form arginosuccinate. Arginosuccinase then catalyzes the splitting of arginosuccinate to fumarate

and arginine. The enzyme arginase completes the cycle by cleaving arginine into urea (for elimination) and ornithine (for recycling). The urea cycle and compounds involved in it are shown in Figures 14-14 and 14-15.

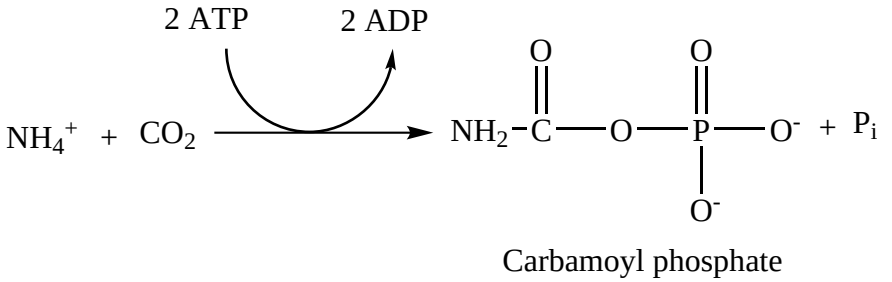


Figure 14-13: Formation of carbamoyl phosphate.

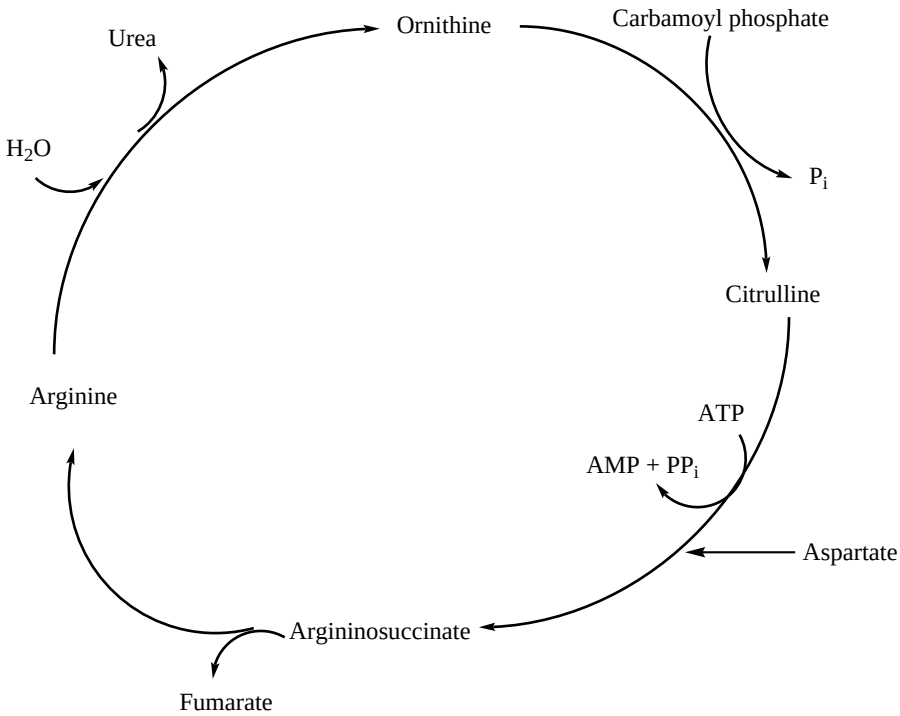
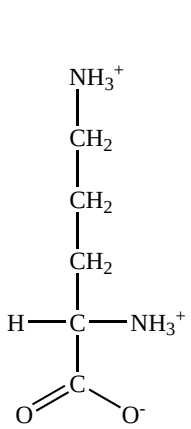
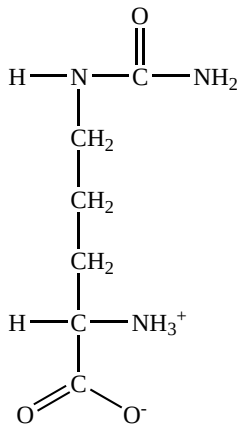


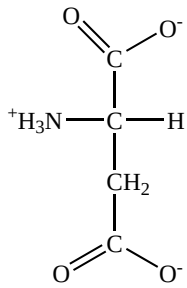
Figure 14-14: Overview of the urea cycle.



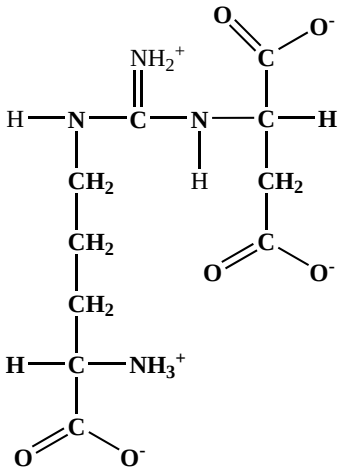
Ornithine



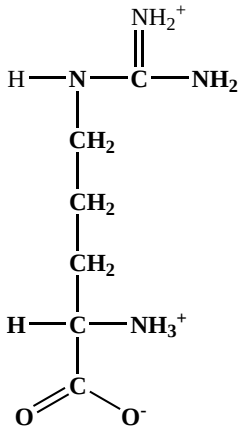
Citrulline



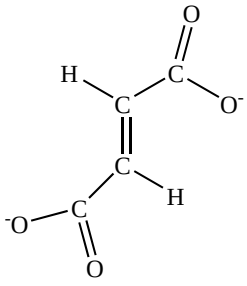
Aspartate



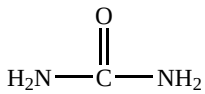
Arginosuccinate



Arginine



Fumarate



Urea

Figure 14-15: Compounds from the urea cycle.

Amino Acids Once Again

The synthesis of proteins requires 20 amino acids. If not readily available, humans can synthesize ten of these amino acids. These are the non-essential amino acids. The remaining ten amino acids, the essential amino acids, must come from the diet. Table 14-3 summarizes these amino acids.

Table 14-3 Essential and Non-Essential Amino Acids

Essential Amino Acids	*Non-Essential Amino Acids*
Arginine*	Alanine
Histidine	Asparagine
Isoleucine	Aspartate
Leucine	Cysteine
Lysine	Glutamate
Methionine	Glutamine
Phenylalanine	Glycine
Threonine	Proline
Tryptophan	Serine
Valine	Tyrosine

** Not essential in adults*

A *complete protein* supplies all essential amino acids. Not all proteins are complete — many are *incomplete proteins*. In order to avoid disorders due to amino acid deficiencies, the human diet must contain complete proteins.

Transamination is important in the biosynthesis of alanine, aspartate, and glutamate. It is easy to convert aspartate to asparagines and glutamate to glutamine. The synthesis of proline requires four steps beginning with glutamate. The synthesis of serine begins with the glycolysis intermediate 3-phosphoglycerate, and after three steps serine forms. It is easy to convert serine to glycine. If sufficient phenylalanine is available, the catalyzed oxidation converts it to tyrosine. If sufficient methionine is available, the body can convert some of the excess to cysteine. Arginine comes from the urea cycle, but infants do not get sufficient quantities from this source.

Metabolic Disorders

When something is out of whack with an organism's metabolism, problems arise that must be treated.

Gout

Gout is the result of overproduction of uric acid, which leads to the precipitation of sodium urate in regions of the body where the temperature is lower than normal (37°C). These low temperature regions are commonly found in the joints of the extremities. Sodium urate may also precipitate as kidney stones. Treatment is partially dietary and partly with drugs. Dietary restrictions include limiting the intake of foods high in nucleic acids (meats) and alcohol, which aggravates the conditions. Doctors often prescribe drugs that inhibit the enzyme that produces uric acid.

Gout may also be the result of faulty carbohydrate metabolism. A deficiency in glucose-6 phosphatase forces phosphorylated carbohydrates to form ribose 5-phosphate instead of glucose. Excess ribose 5-phosphate leads to excess PRPP, which, in turn, stimulates the synthesis of purines. The excess purines cause the production of more uric acid.

Lesch-Nyhan syndrome

Lesch-Nyhan syndrome is another example of defective purine catabolism leading to excess uric acid. Patients with this disorder normally excrete 4–5 times as much uric acid as gout patients do. This is a genetic disease that is a recessive X-linked trait, the trait is carried by the mother and is passed on to her son. There is no treatment for this disease at the present time.

Albinism

Albinism, a recessive trait, is an inborn error of tyrosine metabolism. Tyrosine is the precursor of melanin, the pigment responsible for hair and skin color. In at least one form of albinism, the problem appears to be due a deficiency of the enzyme tyrosinase. A variation of albinism involves a temperature-sensitive form of tyrosinase. The enzyme is only effective at lower than normal temperatures, as found in the extremities. This form of tyrosinase is responsible for the coloration of Siamese cats.

Alkaptonuria

Alkaptonuria is a benign condition that manifests itself as a darkening of the urine. The condition is the result of a problem in the catabolic breakdown of phenylalanine and tyrosine. A defective enzyme leads to an accumulation, and subsequent elimination, of one of the reaction intermediates.

Phenylketonuria

Phenylketonuria, or PKU, is the result of a deficiency in the enzyme phenylalanine 4-monooxygenase, which results in a problem in phenylalanine metabolism. The consequence is an accumulation of phenylalanine in the blood. High levels of phenylalanine enhance transamination to form abnormally high levels of phenylpyruvate. High levels of phenylpyruvate damage the brains of infants with the condition.

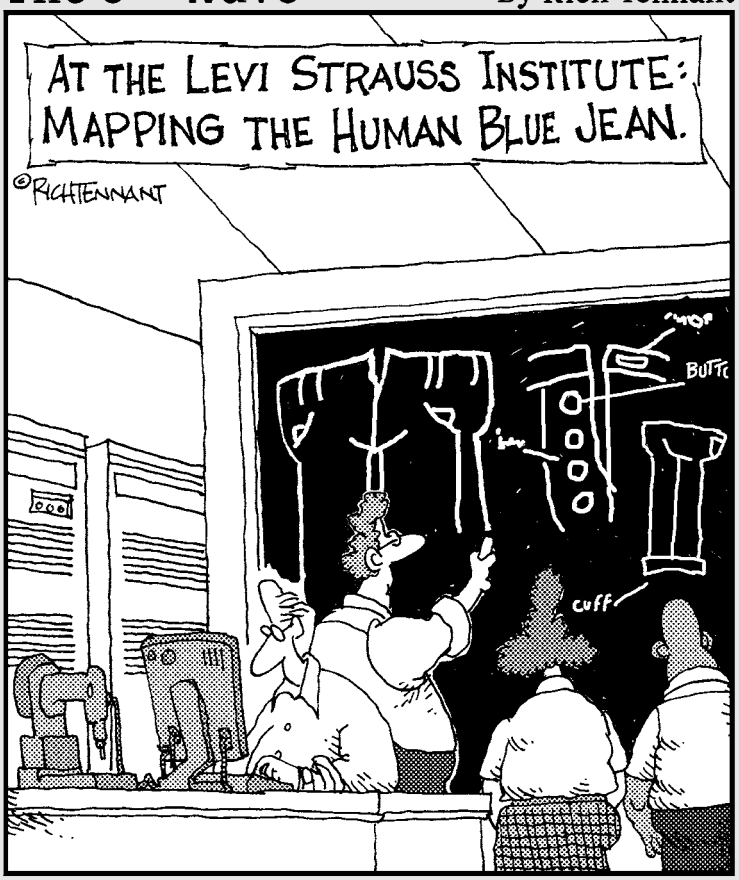
The high levels of phenylalanine lead to competitive inhibition of the enzymes responsible for melanin production from tyrosine. Because little tyrosine converts to melanin, afflicted infants have light blonde hair and fair skin (similar to albinism).

Early diagnosis in infants is important to prevent brain damage. One test for PKU is to add $FeCl_3$ to the patient's urine. Phenylpyruvate reacts with iron ions to produce a green color. Another test is to assay for phenylalanine 4-monooxygenase activity. Treatment consists of maintaining a diet low in phenylalanine until at least the age of three.

Part V

Genetics: Why We Are What We Are

The 5th Wave By Rich Tennant



In this part . . .

We roll up our sleeves and return to the subject of genes and DNA to look at them much more closely. We cover the way DNA replicates itself and look at a number of applications related to DNA sequencing. Then it's off to RNA transcription and protein synthesis and translation. At the very end we mention some of the goals and questions sought by the Human Genome Project.

Chapter 15

Photocopying DNA

In This Chapter

- Learning about replication
- Checking out recombinant DNA
- Examining DNA ssequencing
- Discussing ethical issues

In 1958, Francis Crick postulated what became the "central dogma of molecular biology." In this postulate, he, and later others, reasoned that DNA was the central source of genetic information and that it passed on some of this information to form RNA, which, in turn, passed this information on to form proteins. This central dogma is an extension of the one-gene one-protein hypothesis. To achieve this, the DNA must be able to pass on its information both to later generations (*replication*), and to RNA (*transcription*). RNA must finish the series by forming the appropriate proteins (*translation*).

Some RNA, especially some viral RNA, can undergo replication and even reverse-transcription — thus, RNA can produce both RNA and DNA. Genetic researchers initially thought this was in conflict with the central dogma; but Crick reasoned that RNA creating DNA was an extension of this postulate.

Many of the viruses capable of reverse-transcription are cancer causing.

The primary structure of DNA consists of two polynucleotide strands held together by hydrogen bonds. Adenine forms hydrogen bonds to thymine, and cytosine forms hydrogen bonds to guanine (Figure 15-1). The sequence of nitrogen bases contains the genetic information. The DNA molecules wrap around a protein called a *histone* — the combination of eight histones with the associated DNA is a *nucleosome.* (We talk more about histones in Chapter 16.)

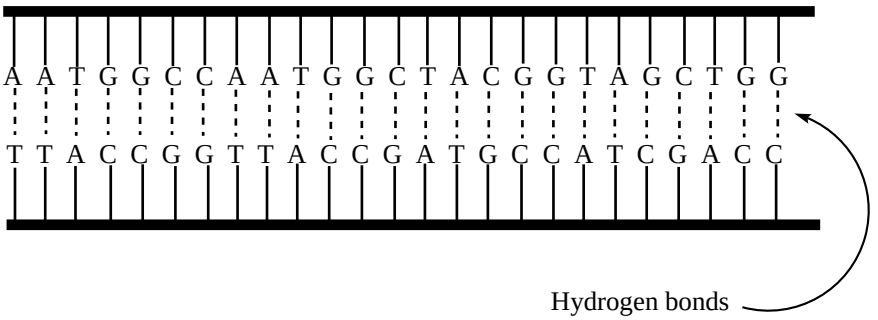


Figure 15-1: A schematic illustration of the base pairs present in a segment of DNA.

A *gene* is a portion of a DNA molecule that carries specific information. The portion of the gene coding for that specific information is called an *exon*. The portion of a gene that does not code for specific information is an *intron*.

Let's Do It Again: Replication

Replication is the process that produces new DNA molecules. One DNA molecule produce two DNA molecules in a process where the DNA must unwind and open — kind of like a zipper. New nucleotides bind to the backbone of each strand of the opened DNA by forming hydrogen bonds to the nucleotides (the zipper's "teeth") that are already present. The process proceeds along the opening DNA strand until each half of the original DNA has a complementary strand hydrogen bonded to it. The result is two DNA double helices each with half old DNA, and half new. It doesn't sound like much fun, but it works for DNA. Replication is illustrated in Figure 15-2.

The specific hydrogen bonding forces the new strands to contain a nucleotide sequence that is complementary to the nucleotide sequence in the old strand. Therefore, it can create an exact duplicate of the original DNA.

This description of replication is a simplification. It barely scratches the surface of this complicated process, but it should give you enough background information in order to understand what comes next.

The first step in understanding replication was the discovery of DNA polymerase from *Escherichia coli*. Subsequent studies showed that this enzyme needed a DNA template and all four deoxyriboside triphosphates (dATP, dCTP, dGTP, and dTTP). In addition, a short section of RNA called a *primer* is also needed. The enzyme prefers a single DNA strand for the template in order to produce a complementary strand.

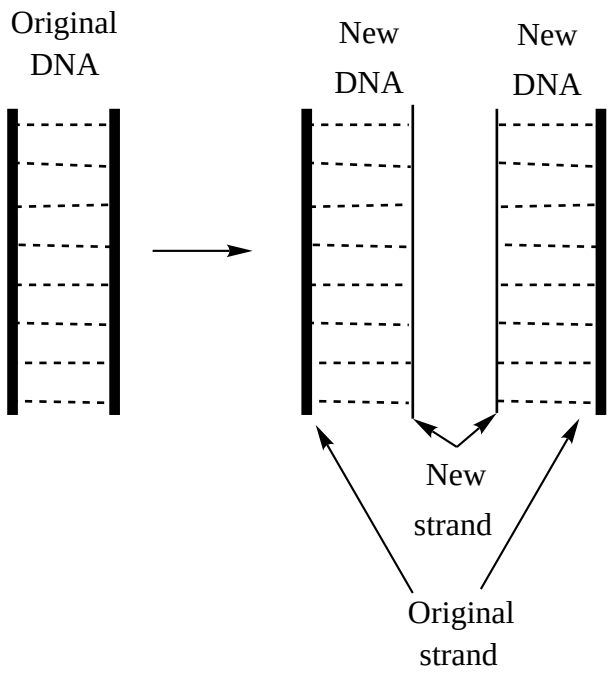


Figure 15-2: A simplified representation of replication.

During replication, simultaneous duplication of the two strands of DNA occurs. Because the two strands of DNA are anti-parallel, the mode of synthesis is different for each strand, but the overall process is the same: moving from one end to the other. For one strand the synthesis is from 5' → 3'. On the other strand it appears to be from 3' → 5', but in actuality it is also 5' → 3'. There is a complication on the 3' → 5' strand (which we discuss later in this section). See Figure 15-3.

The initiation of replication begins at a particular site, and, once initiated, a series of fragments form discontinuously along one strand and continuously along the other strand. These discontinuous fragments, known as *Okazaki fragments,* contain from 1,000 to 2,000 nucleotides. The synthesis of the fragments is always in the 5' → 3' direction. See Figure 15-4. Note that Figures 15-3 and 15-4 appear to be different at first glance. Figure 15-3 represents a simplified view of the overall process, whereas Figure 15-4 illustrates in more detail how this overall process occurs.

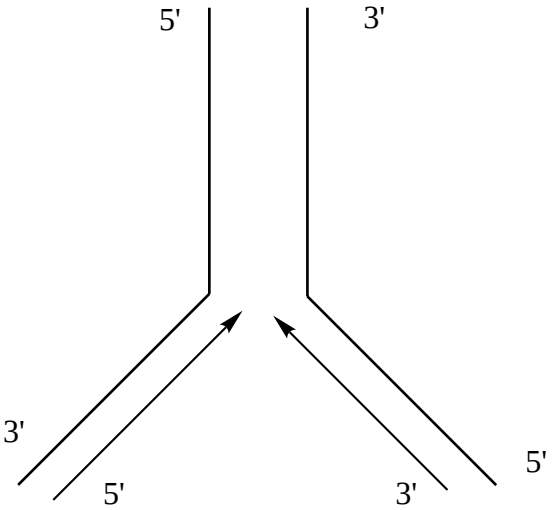


Figure 15-3: A simplified scheme of the replication of DNA.

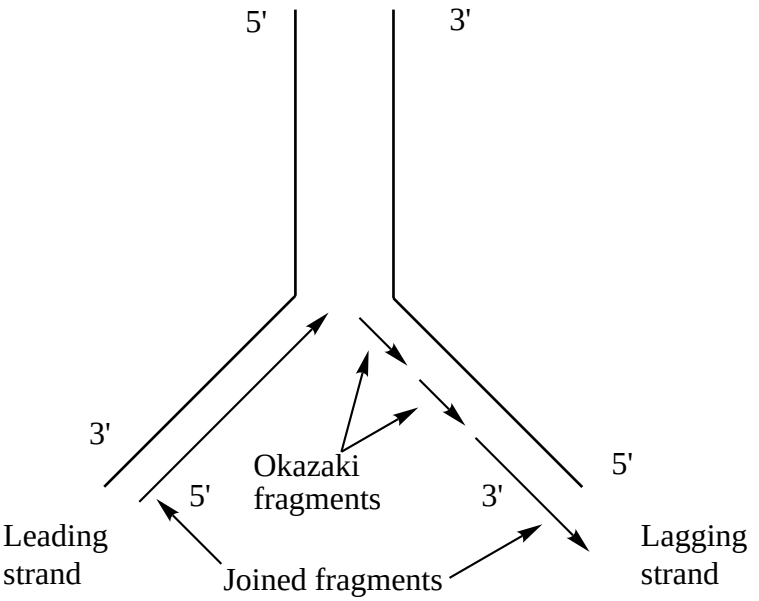


Figure 15-4: Detailed scheme of the replication of DNA.

Researchers unexpected found that RNA synthesis is a prerequisite for the replication of DNA. Initially, an RNA primer, typically 20 to 30 nucleotides in length, forms on a single DNA strand. Once formed, deoxyribonucleotide nucleotides add to the 3' terminus. Later, it is necessary to remove the RNA primer and attach the appropriate DNA fragment to produce the completed DNA.

At least a portion of the double-stranded DNA must be separated before replication can occur, and the separated portions can serve as templates. Enzymes known as *helicases* are responsible for this separation. The energy needed comes from the hydrolysis of ATP. The mechanism of separation is not well understood and is still under investigation. Apparently, the helicase binds more strongly to one strand of the DNA than the other so that the enzyme squeezes in and pushes the other strand away. ATP hydrolysis provides the energy necessary to cause the enzyme to move along the one strand nucleotide by nucleotide. This results in regions of the DNA opening like the afore-mentioned zipper.

DNA polymerases

DNA polymerases are the enzymes responsible for joining the nucleotide triphosphate fragments to produce a strand of DNA, acting as the bricklayers and carpenters in its construction. This process will only occur in the presence of a DNA template (parent DNA). Before the enzyme can connect a nucleotide, the nucleotide must bind to the appropriate site on the template.

There may be more than one DNA polymerase present in a cell. For example, in *E. coli* three different enzymes perform this task. These enzymes may also act as exonucleases. An *exonuclease* has the opposite function as a polymerase; that is, it removes nucleotides from the DNA strand.

The addition of the nucleotides is always to the 3' end of a polynucleotide chain. DNA polymerases cannot start building a nucleotide from scratch — there must be a polynucleotide already present. In contrast, RNA polymerase *can* begin from scratch. RNA polymerase generates the RNA primer, using ribonucleotides, at the beginning of replication. DNA polymerase then takes over the task and adds deoxynucleotides to the RNA primer. The polymerization requires the presence of two metal ions to enable the joining of the nucleotide to the polynucleotide.

Replication of DNA needs to be error free to ensure proper transmission of genetic information, and DNA polymerases are extremely effective in reducing errors. The enzyme binds tightly to the template and to the incoming nucleotide. This nucleotide is initially bound to the template through hydrogen bonding. If the wrong nucleotide is present, the subsequent binding to the polymerase is ineffective, and the nucleotide is "rejected." In addition to this checking, DNA polymerase also proofreads the preceding nucleotide to make sure it is correct. If the wrong nucleotide is present, it does not fit properly, making it necessary to remove the erroneous nucleotide from the polynucleotide so that the correct nucleotide may enter. The exonuclease portion of the polymerases performs this function. The polymerase proofreads the polynucleotide chain as polymerization proceeds. Proofreading is in the reverse direction (3' $\rightarrow$ 5'). There must be a nucleotide already in place before the polymerase can proofread. (We hope that our proofreader is as good as the DNA polymerases.)

The current model of DNA replication

In vitro studies show that in *E. coli,* replication begins when a protein binds a region of the DNA containing four specific binding sites. This is the *origin of replication* site. Once this protein binds, a helicase enzyme attacks the DNA and begins to unwind and separate the two strands. A third protein enters and holds the DNA strands open so that replication can continue. This third protein is the single-strand binding protein. The partially opened DNA and associated proteins are called the *prepriming complex* (Figure 15-5).

It is necessary to expose the DNA templates in this manner. A DNA strand may have more than one origin of replication site — this allows replication to occur in many places at one time. Simultaneous replication allows the cell to replicate the entire strand in less time.

Replication cannot continue until the exposed template is primed. A type of RNA polymerase known as *primase* binds to the prepriming complex in a region known as the *primosome.* Primase synthesizes a short RNA segment of about five nucleotides. Primase is capable of performing this function because its proofreading ability is not as efficient as that of DNA polymerase. For this reason, a nucleotide doesn't need to already be present to be checked. Because the primer consists of ribonucleotides instead of deoxyribonucleotides, it is temporary and will be detected and removed later. Once removed, the appropriate deoxyribonucleotides join to complete the DNA strand (Figure 15-6).

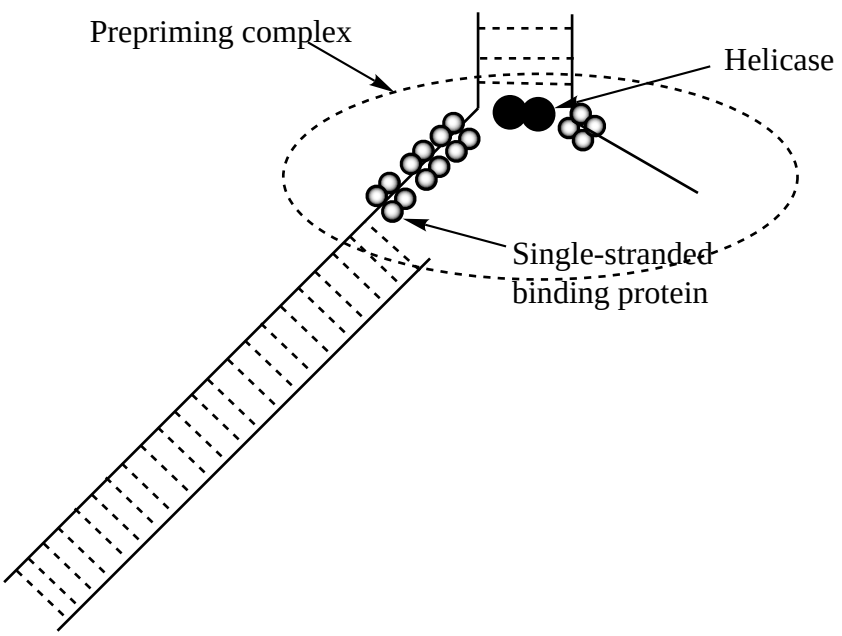


Figure 15-5: A simplified view of the prepriming complex.

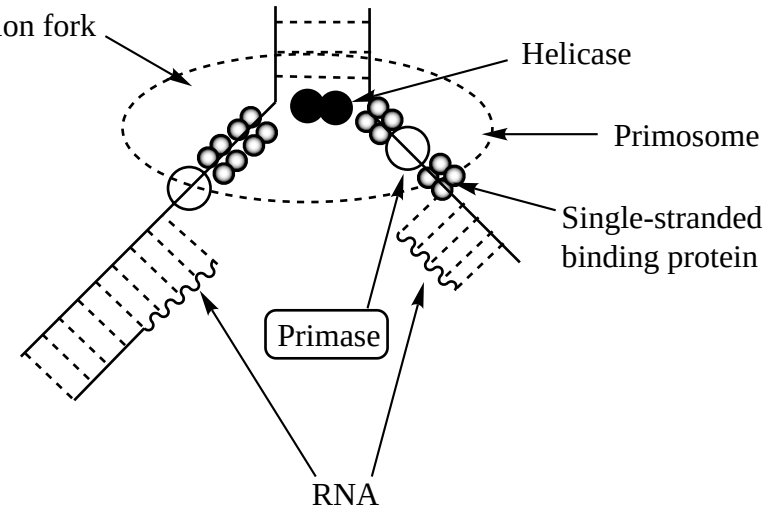


Figure 15-6: Formation of the RNA primer.

Although both strands of DNA serve as templates, the replication process differs on each strand. The point where the strands split and replication occurs is the *replication fork.* Because the two strands are anti-parallel, and DNA polymerase only works in the 5' → 3' direction, direct replication only works on one strand — called the *leading strand.* The other strand is the *lagging strand.*

As the DNA strands separate, eventually there is enough room to begin synthesis in the reverse direction on the lagging strand. (The reverse direction on the anti-parallel lagging strand is still 5' → 3'.) Replication on the lagging strand is discontinuous, and fragments of about 1,000 nucleotides form, called, as we have already noted, Okazaki fragments. DNA ligase then joins the fragments to produce a continuous strand.

DNA polymerase III holoenzyme (complete enzyme) simultaneously produces DNA on both the leading and lagging strands, though the mechanisms on the two strands are different. On the leading strand the process is continuous, whereas on the lagging strand it is discontinuous and more complex. To carry out the polymerization on the lagging strand, this strand loops around so that polymerization in the 5' → 3' direction can take place. After about 1,000 nucleotides — an Okazaki fragment — the polymerase releases the loop and begins a new loop and fragment. Each Okazaki fragment has a RNA primer. DNA polymerase I synthesizes DNA in the gaps between the fragments and removes the primer section. DNA ligase then joins the fragments (Figure 15-7). Wow! John wishes the carpenters who built his new house were that efficient!

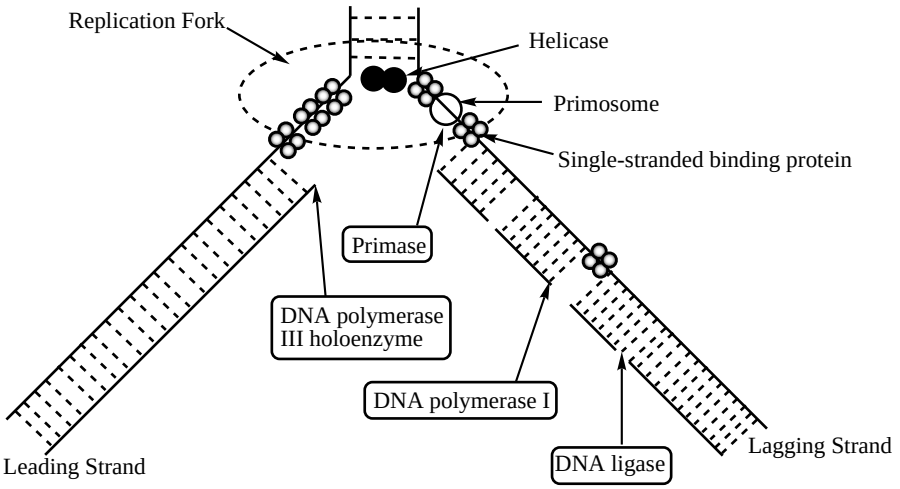


Figure 15-7: An expanded representation of the replication fork.

The ends of the DNA strands require a different procedure than does the majority of the strand, and this procedure is especially important on the lagging strand. If care were not taken, each replication cycle would result in a shorter DNA strand, eventually leading to the loss of important genetic material. To resolve this problem the ends of the DNA strands contain telomeres. *Telomeres* are DNA segments containing hundreds of repeating units. In humans, the repeating units are the hexanucleotide AGGGTT. The enzyme *telomerase,* in humans, detects the primer sequence GGTT and repeatedly attaches the hexanucleotide units, completing the DNA strand.

Mechanisms of DNA repair

All cells have a variety of DNA repair mechanisms, which are necessary to repair defective DNA and ensure retention of genetic information. Damage to DNA may occur during replication or by the action of radiation or chemicals. There is a rare error known as *xeroderma pigmentosum,* which impairs these repair mechanisms. Individuals suffering from it are extremely susceptible to cancers, especially skin cancers. Eventually the skin cancers metastasize, leading to death. The three general types of repair mechanisms are

- Direct repair
- Base-excision repair
- Nucleotide-excision repair

One example of damage needing repair is the formation of a thymine dimer (Figure 15-8) by ultraviolet (UV) light. The *thymine dimer* is an example of a

pyrimidine dimer, and its presence causes distortion of the DNA in the region. Other problems include base mismatches and missing or additional bases.

Direct repair

Here, the correction of the problem occurs in place. The photoreactivating enzyme, DNA photolyase, binds to the cyclobutane ring present in a thymine dimer, using light energy to cleave this dimer into the original bases.

Base-excision repair

In base-excision repair, the correction of the problem involves removal and replacement of the base. This is necessary whenever a modified base is present. There are various causes of modified bases, such as radiation or certain chemicals. The presence of a modified base normally results in a recognizable distortion in the DNA molecule. An enzyme, behaving as a glycosylase, cleaves the glycosidic bond to release the base from the deoxyribose. The result is an AP site (AP meaning *apurinic* or *apyrimidinic*). With apurinic, the purine base is absent, in apyrimidinic the pyrimidine base is absent. An AP endonuclease recognizes this site and cuts the DNA backbone adjacent to the site. Next, a deoxyribose phosphodiesterase completes the removal of the remaining deoxyribose phosphate. DNA polymerase I then inserts a replacement nucleotide to match the nucleotide in the complementary DNA strand. Finally, DNA ligase connects the units to yield the repaired strand — kind of like an electrician cutting out a bad circuit and splicing a good one in its place.

Nucleotide-excision repair

In nucleotide-excision repair, the correction of the problem involves the removal of a segment of DNA around the problem followed by its replacement. When this mechanism occurs, a DNA strand on both sides of the error is cut from the DNA strand. Typically, an exonuclease removes a 12-nucleotide section. DNA polymerase I then synthesizes a replacement segment of the strand. DNA ligase then finishes the repair.

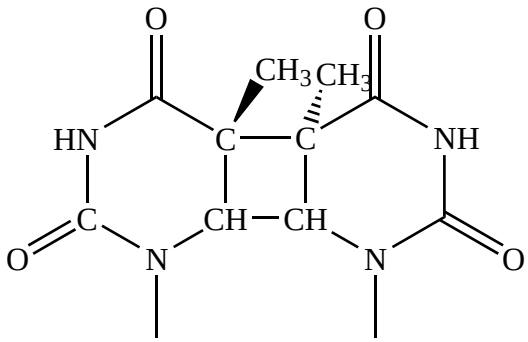
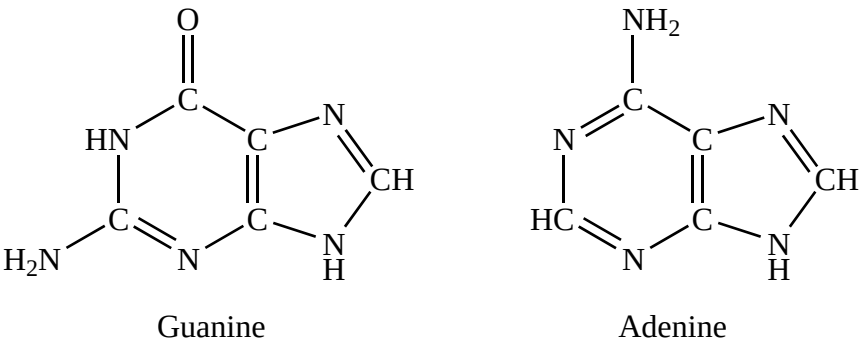


Figure 15-8: Structure of a thymine dimer.

Mutation: The good, the bad, and the ugly

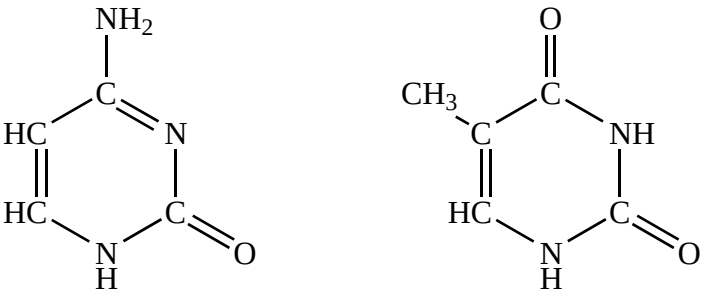
Several types of mutations are known. DNA repair mechanisms try to prevent new mutations — however, such mechanisms are not always effective. Known mutations include the substitution of one base pair for another, the insertion of one or more base pairs, and the deletion of one or more base pairs. Changes, especially subtle ones, may occur during or after replication.

The substitution of one base for another is a common mutation. There are two types of substitutions. The replacement may be of a purine by the other purine (Figure 15-9) or the replacement of a pyrimidine by the other pyrimidine (Figure 15-10). This type of error is a *transition.* The other type of substitution is the replacement of a purine for a pyrimidine or vice versa. This latter type is a *transversion.*



Guanine Adenine

Figure 15-9: The purines.



Cytosine Thymine

Figure 15-10: The pyrimidines.

Any uncorrected discrepancy in the genetic code will become "normal" in all future generations. The new genetic code is a *mutation.* The change in the base sequence may or may not affect the amino acid for which the codon codes. For example, changing from GTT (coding for leucine) to GTG (also coding for leucine) results in no change. However, if the change results in coding for a different amino acid, the resultant protein will function differently. If the new protein exhibits improved function, the organism benefits from the change. But if the new protein exhibits impaired function — the more likely situation — the organism suffers from the change. Problems from impaired function are genetic diseases. Table 15-1 lists some of these.

Table 15-1 Some Genetic Diseases in Humans

Disease	*Defective Protein*
Acatalasia	Catalase
Albinism	Tyrosinase
Cystic fibrosis	CF transmembrane conductance regulator
Fabray's Disease	α-Galactosidase
Gaucher's Disease	Glucocerebrosidase
Goiter	Iodotyrosine dehalogenase
Hemochromatosis	Hemochromatosis
Hemophilia	Antihemophilic factor (factor VIII)
Hyperammonemia	Ornothine transcarbamylase
McArdle's Syndrome	Muscle phosphorylase
Niemann-Pick Disease	Sphingomyelinase
Phenylketonuria	Phenylalanine hydroxylase
Pulmonary emphysema	α-Globulin of blood
Sickle cell anemia	Hemoglobin
Tay-Sachs Disease	Hexosaminidase A
Wilson's Disease	Ceruloplasmin (blood protein)

Restriction enzymes

Although not directly related to replication, restriction enzymes are important tools in genetic research. *Restriction enzymes,* or *restriction nucleases,* are capable of cutting DNA into fragments. These were first found in prokaryote cells like *E. coli* where these enzymes locate and destroy invading DNA, such as that of a bacteriophage, but leave the cell's own DNA alone. Recent research focuses on the fact that it is possible to manipulate these fragments so that DNA ligases can join the fragments into new DNA. Restriction enzymes are important in vitro biochemical tools that act as very accurate molecular scalpels. Cleavage may leave both DNA strands of equal length or one strand may be longer than the other (a *staggered* cut).

More than 100 restriction enzymes have been identified and are available for research. These enzymes recognize specific regions in the DNA and cleave DNA molecules into specific fragments. Because these fragments are smaller than the parent DNA is, they are easier to manipulate and analyze. Testing a strand of DNA with a series of restriction enzymes can provide a fingerprint of cleaved fragments. In fact, you can map the structure of DNA.

Many times in reading descriptions of genetic determination and modification, you will run across the terms *in vivo* and *in vitro. In vivo* means in the cell, whereas *in vitro* means in a test tube.

Mendel Rolling Over: Recombinant DNA

Recombinant DNA technology allows the synthesis of DNA strands that contain one or more genes not originally present. The addition of new genes enables an organism to produce new biochemicals. For example, *E. coli* has been engineered to produce human insulin. Recombinant DNA technology also allows biochemists to add a gene to compensate for a defective gene.

Restriction enzymes are capable of removing DNA fragments of interest. It is then necessary to join one of these fragments to another DNA strand for replication. The DNA to which the fragment of interest is attached is the *vector.* Common vectors include plasmids. A *plasmid* is a naturally occurring DNA circle. The first step in adding the fragment is to create a staggered cut in the DNA of the vector. The longer end of the staggered cut is a "sticky" or cohesive end. It is possible to attach any DNA fragment to the sticky end if it has the complementary DNA sequence. The complementary sticky end will be present if the same restriction enzyme was used to excise the fragment of interest. DNA ligase completes the joining of the fragment to the vector (Figure 15-11).

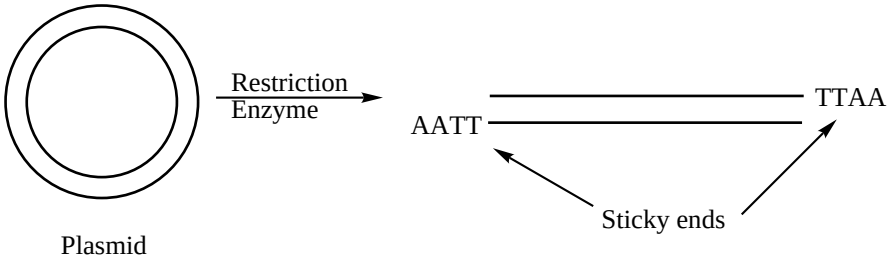


Figure 15-11: Opening of a plasmid by a restriction enzyme such as Eco RI.

It is possible to bond a DNA linker to a DNA molecule to make it susceptible to a particular restriction enzyme. By this method, the cohesive ends characteristic of any restriction enzyme may be added to almost any DNA molecule. The completed DNA can undergo replication.

Plasmids are, to a certain extent, accessory chromosomes. They can replicate independently of the host chromosomes. Thus, there may be multiple copies of a particular plasmid within a cell. This replication, in general, makes plasmids more useful as vectors than host chromosomes. Thus far, these plasmids have only been shown to be relevant in bacterial organisms.

The addition of "new" genes to an organism produces an organism that may be considered a new species. There is somewhat of a risk that these organisms could infect humans and lead to a new disease for which there is no known treatment. To minimize potential risks posed by these organisms, researchers either use enfeebled (weakened) organisms or ones that do not infect humans.

Patterns: Determining DNA Sequences

Restriction enzymes are a major tool in the determination of the base sequence in DNA. The cleaved DNA fragments are significantly smaller than the parent DNA is, making manipulation and analysis significantly easier. To separate the fragments after cleavage, gel electrophoresis is often used.

Gel electrophoresis is a biochemical technique used to separate and purify proteins and nucleic acids that differ in charge, size, or confirmation. The sample is placed into wells within a gel — a polymer that is specifically formulated for the type of analysis. This gel is in the shape of a thin slab. When separating proteins or small nucleic acids (DNA, RNA, and so on) cross-linked polyacrylamide is used. For separating larger nucleic acids, agarose, an extract from seaweed, is used. These gels have the consistency of Jell-O, but probably don't taste nearly as good.

The gel is immersed in a buffer solution, and an electrical current is applied to the ends of the gel. The charged species within the sample migrate toward one or the other of the electrodes. Proteins may have either a positive or a negative charge, but, at the proper pH, nucleic acids have only a negative charge. The positively charged species move toward the negatively charged end of the gel, and the negatively charged species move toward the positively charged end. Normally a buffer adjusts the pH so that all the species of interest have either a positive or negative charge (Figure 15-12).

Different molecules move at different speeds through the gel. When the smaller, faster molecules have about reached the end of it, the process is stopped, and the molecules are stained to make them visible. Sometimes, agents are added to cause the molecules to fluoresce (glow) under UV light. Then a photograph of the gel may be taken as it is exposed to the UV light. When several samples, including a known sample, are run side by side, the molecular weight of a sample component may be determined. This is one step in the identification of unknown components.

The separation of DNA fragments by gel electrophoresis readily distinguishes even minor differences between the fragments. Different gels are useful in separating large fragments than are useful in the separation of small fragments. In some gels, it is possible to distinguish between fragments differing by one base in several hundred. Modification of the electrophoresis method provides further separation. Each type of DNA gives a different pattern, making it possible to distinguish between two different samples. Two samples giving identical patterns must be from the same source or from identical twins.

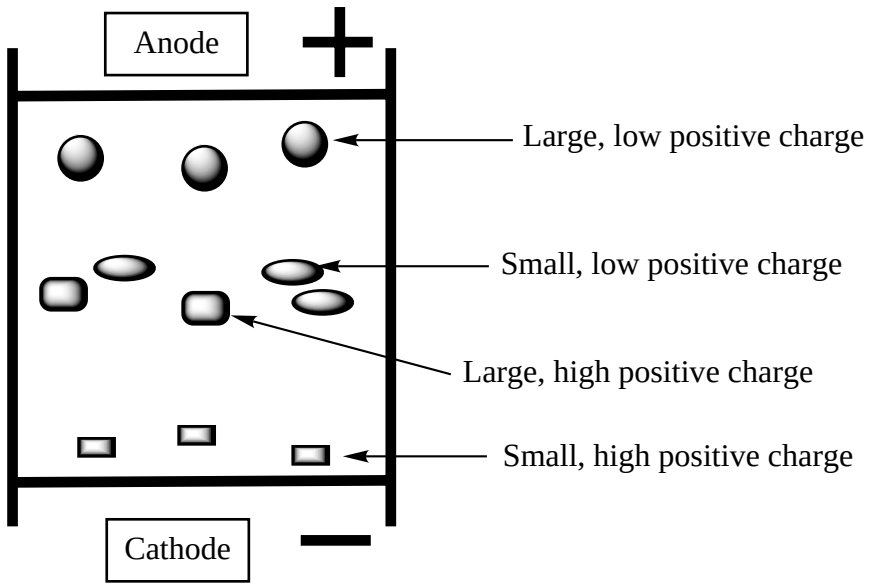


Figure 15-12: Gel electrophoresis.

In the analysis and manipulation of genetic material, it is advantageous to be able to identify whether a certain sequence of nucleotides is present. The general method for finding a particular sequence of nucleotides in DNA was developed by Edwin Southern. It is called *Southern blotting.* This method uses radioactive ^{32}P as a label that is easily detectible. This radioisotope is incorporated into the phosphate in some of the nucleotides. Determination of a particular nucleotide sequence in RNA is achieved through *Northern blotting,* and protein identification through *Western blotting.* (The names Northern and Western do not refer to persons with that name, but are by analogy to Southern. We don't know what happened to Eastern.) Alternatively, Southern, Northern, and Western blotting are DNA, RNA, and protein blots, respectively.

Determining the base sequence

Since the first isolation of DNA, a number of methods have been developed to determine the base sequence. In general, the *Sanger dideoxy method* has replaced all others. It employs the controlled termination of replication with modified nucleotides containing dideoxyribose in place of deoxyribose.

DNA fragments produced by employing restriction enzymes are denatured to give single-stranded DNA. (*Denaturing* typically involves heating a DNA-containing solution to 96°C for a few seconds.) Four samples of this DNA are treated separately to produce double-stranded DNA through replication, with each sample containing a small quantity of a different dideoxy nucleotide. The dideoxy nucleotide contains dideoxyribose (Figure 15-13). The absence of an additional oxygen atom in dideoxyribose means that there is no 3' hydroxyl group available to continue replication. Thus, the incorporation of a dideoxynucleotide terminates the DNA chain.

One of the four samples will contain a small quantity of the dideoxy analog of the nucleotide dGTP. This "defective" unit enters the new DNA strand as the complement to a cytosine base in the original fragment. Separation of the new material from the original strand material gives a set of DNA fragments of varying length. These fragments are then separated by electrophoresis according to length (size). The length of each of these fragments locates the position of each C in the original strand. The other three samples give the positions of all A, T, and G bases in the original strand.

Fluorescence tagging is a useful modification to this method. Each of the dideoxy nucleotides has a different fluorescent tag attached. After attaching the tags, it is possible to conduct all four experiments in one container. Separating the fragments by electrophoresis and examining the tags gives a colored pattern showing all the bases in sequence. This method works for fragments of up to 500 bases.

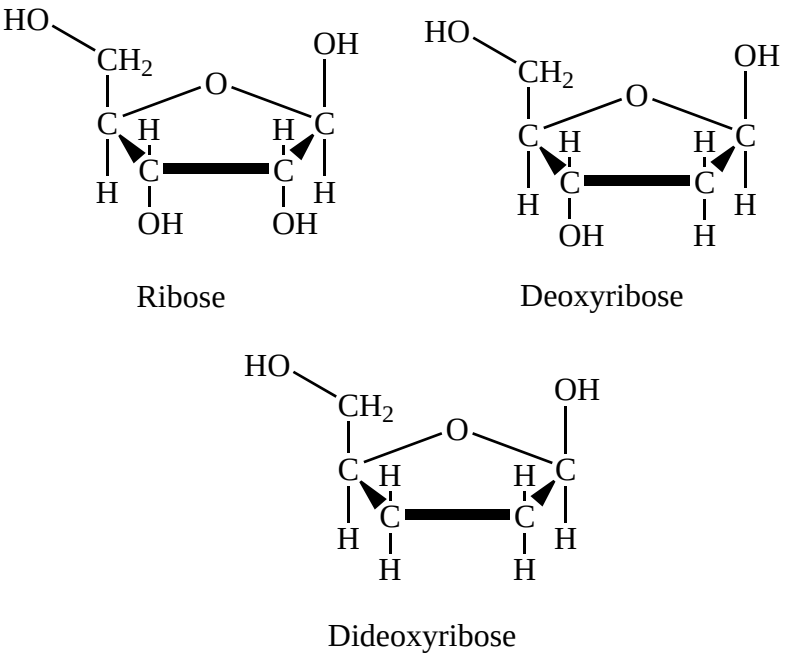


Figure 15-13: Structures of ribose, deoxyribose, and dideoxy-ribose.

To do the studies described, you need a sufficient amount of genetic material. Lack of sufficient quantities of sample has been a problem, especially with forensic evidence. Therefore, ways of quickly duplicating sufficient quantities of identical DNA fragments or producing a number of DNA strands from a very small sample were developed. *Polymerase chain reaction* (PCR), is a useful method to amplify specific DNA sequences. It is an in vitro procedure where it is necessary to know the base sequences, the flanking sequences, adjacent to a particular target sequence. However, it is not necessary to know the base sequence in the target region. Denaturation of a DNA sample provides two separate strands. Two primers are added to the mixture and one primer will attach to the flanking sequence of each strand. DNA polymerase begins replication starting at each of these primers. Repeating these steps quickly generates a large quantity of DNA. After 30 or so cycles, a billion-fold amplification occurs. Thirty cycles take less than one hour.

The butler did it: Forensic applications

Scientists can identify a species by the isolation and examination of the DNA sequences unique to that species. For example, DNA analysis is useful in the identification of organisms, such as bacteria, that may be polluting our water, food, and other samples. It has been used to establish pedigrees for livestock

breeds as well as identify endangered species in the prosecution of poachers. However, the application that has received the most publicity is in the area of *forensics*.

Because an individual's DNA comes from both the mother and father, it is unique to that individual (except in the case of identical twins). Even brothers and sisters, including fraternal twins, with the same parents show some variation in their DNA. This fact makes DNA analysis very valuable in forensics investigations (as anyone who ever watched an episode of *CSI* can attest).

In order to identify an individual, forensic investigators examine 13 regions (markers) of the DNA sample that vary significantly from individual to individual. There is a very small chance that two individuals might have the same DNA pattern at these 13 regions, but it is only about one chance in a billion. The investigation of additional markers can improve the discriminating ability of the procedure. Investigators then combine the results into a DNA profile — also known as a DNA fingerprint — of the individual.

You can isolate DNA samples from blood, hair, bone, fingernails, teeth, and any type of bodily fluid. In a typical crime scene analysis, samples are taken from the evidence and suspects; the DNA is extracted and then analyzed for the specific markers. A match of a single marker does not prove that an individual was at the crime scene, but the matching of four or five markers indicates a very high probability that the individual was present. PCR may be necessary if the sample is very small (see preceding section).

Methods of analysis

Several techniques are used in DNA analysis. The three most common are RFLP, PCR, and STR. In RFLP (*Restriction fragment length polymorphism*), the DNA sample is digested with a specific enzyme, a restriction endonuclease. This enzyme cuts DNA at a specific sequence pattern. The presence or absence of these sites in a DNA sample leads to variable lengths of DNA fragments. Gel electrophoresis then separates these fragments.

RFLP was one of the original forensic DNA analysis techniques. However, it requires relatively large amounts of DNA and samples contaminated with dirt and mold are difficult to analyze with RFLP. It has been somewhat replaced with polymerase chain reaction (PCR) enhancement, followed by STR analysis.

PCR (discussed in the earlier section "Determining the base sequence") is a useful technique that reduces the sample size requirement of RFLP — in essence it is a DNA amplifier. PCR quickly makes millions of exact copies of the DNA sample. Using PCR, DNA analysis can be done on a sample as small as a few cells and on samples that are extensively degraded. After PCR treatment, it is possible to analyze the sample with RFLP or STR.

In STR (*Short tandem repeat*) analysis, the DNA sample is quickly examined for 13 specific regions. The FBI uses this standard STR profile in its CODIS (Combined DNA Index System) program, which links national, state, and local databases of DNA profiles from felons, missing persons, and unsolved crime scenes. CODIS has an index of more than 3 million DNA profiles.

Paternity testing

Along with crime scene analysis, paternity testing is one of the most widely used applications of DNA testing. The procedure begins with the collection of DNA samples from the mother, child, and alleged father(s). The DNA profiles of the child and mother are first determined. The markers not inherited from the mother must have come from the biological father. The alleged father's DNA profile is then compared to the child. If the man's DNA profile contains markers common to the child but not the mother, then the probability that he is the biological father is great. Figure 15-14 indicates that Alleged Father 2 is more likely to be the biological father than Alleged Father 1.

Genetic Diseases and Other DNA Testing Applications

DNA testing always seems to find new ways of being useful. It has been used for a number of years, for example, in determining the gender of athletes. In addition to gender testing, the NFL used a strand of synthetic DNA to mark all the Super Bowl XXXIV footballs as a way to combat fraud associated with sports memorabilia. In a different situation, a section of DNA was added to the ink used to imprint all official goods marketed at the 2000 Summer Olympics Games. This same technology is used to tag original artwork, in addition to sports souvenirs.

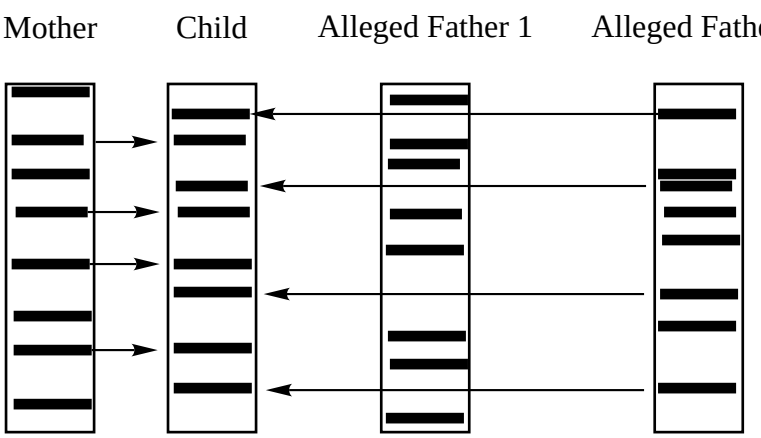


Figure 15-14: Comparison of results for paternity testing.

Genetic diseases are the result of an abnormal pattern in the DNA of an individual. These diseases are inherited, though some individuals are only carriers and not sufferers. Recently quite a bit of research has been done in determining the genetic pattern that is causing the disease, and ways to detect the probability of passing on the disease to offspring. However, methods of treatment for most all of these disease are limited. It is the dream of researchers to find the means of correcting these genetic diseases through genetic modifications. Researchers have investigated several of these genetic diseases in detail. In this section we briefly examine a few of the more well-known genetic diseases.

Sickle cell anemia

Sickle cell anemia is an inherited genetic disease of the blood's *hemoglobin,* a component of red blood cells. Sickle cell anemia is the result of the change of a single amino acid in the protein sequence of hemoglobin. This change involves the substitution of valine (non-polar) for glutamic acid (polar). The condition affects millions of individuals throughout the world, especially those whose ancestors came from Africa, South America, Cuba, Saudi Arabia, and a few other countries. In the United States, it affects about 72,000 people. Sickle cell happens in about 1 in 500 African American births and about 1 in 1,200 Hispanic American births.

Hemoglobin is responsible for carrying oxygen from the lungs to the cells. In an individual with sickle cell anemia, the defective hemoglobin molecules clump together, causing the red blood cells to assume a sickle shape, hence the name. These abnormal cells have trouble squeezing through small blood vessels, causing oxygen depletion in organs and extremities along with episodes of pain. These sickle cells also have a much shorter lifetime in the body, leaving the individual with chronic anemia. Many states now test newborns for sickle cell disease.

Hemochromatosis

Hemochromatosis, one of the most common genetic diseases in the United States, is an inherited disease that causes the body to absorb and store far too much iron. This excess iron is stored in organs, such as the liver, pancreas, and skin (yes, the skin is considered an organ!). It is due to a mutation in the HFE gene, the gene that regulates the absorption of iron from food. If this defective gene is inherited from both parents, then the person will develop hemochromatosis. If the individual inherits the mutated gene from only one parent, the person will be a carrier but will not necessarily develop the disease. About 5 Caucasian people in 1,000 carry both mutated genes, and 1 in 10 is a carrier. Genetic testing can detect it about 90 percent of the time.

Cystic fibrosis

Cystic fibrosis is a chronic and normally fatal genetic disease affecting the body's mucus glands. It targets the digestive and respiratory systems. About 55,000 individuals worldwide have cystic fibrosis. Most of these individuals are Caucasians who have ancestors who came from northern Europe. For the disease to appear it is necessary to inherit the mutated gene responsible for cystic fibrosis from both parents. Estimates are that 1 in 20 Americans carry the abnormal gene. Most of these individuals are not aware that they are carriers. Genetic testing is only about 80 percent accurate.

Hemophilia

Hemophilia is a genetic disorder caused by the lack of the blood-clotting factor stemming from a defective gene on the X chromosome. Females have two X chromosomes, so if there is a defective gene on one, there is little chance that the other one is also defective. However, she will be a carrier. Males, however, only have one X chromosome, so if it is defective, then the individual will develop hemophilia. If a woman is a carrier, she will have a 50 percent chance that her sons will have hemophilia and a 50 percent chance that her daughters will be carriers. Daughters of a hemophilic male will be carriers. Genetic testing can detect the presence of the abnormal gene.

Tay-Sachs

Tay-Sachs is an inherited disease in which a fatty-acid derivative, a lipid called *ganglioside,* accumulates in the brain — the result of a mutation of a specific gene. Although found primarily in the Jewish population, some French Canadians and Louisiana Cajuns also carry the abnormal gene. The symptoms most commonly appear in infants. Death normally occurs before the age of five. Although Tay-Sachs is a very rare disease, it was one of the first genetic diseases for which extensive and inexpensive genetic screening was developed. Screening tests were developed in the 1970s, and Israel offered free genetic screening and counseling. Because of this aggressive testing and counseling, the disease has been almost totally eradicated from Jewish families worldwide.

Ethics of genetic modification and testing

The emerging field of *bioengineering* has raised many ethical questions. One has only to listen to the debates over stem-cell research, gender selection of children, genetic modification to enhance certain traits such as athletic ability, and so on. Public policy decisions related to cost are also being debated, as genetic modification and screening are generally expensive. Is it to be available to only those who can pay, or should there be equal access? Many gray areas concerning genetic modification exist in the field of patent law. There are many questions and concerns but no quick answers.

Although the success in eradicating Tay-Sachs is directly related to genetic testing, such testing is not without its ethical questions. The major concern is one of privacy. DNA samples and profiles can be used to determine parentage and susceptibility to certain genetic diseases. Many people fear that the government, insurance companies, employers, banks, schools, and other organizations could use such information for genetic discrimination. In fact, in the United Kingdom, a man was denied treatment for hemochromatosis because his insurance company claimed it was a preexisting condition. Individuals applying for life insurance have reported other cases of genetic discrimination. Who gets to request the genetic screening and who has access to the results? These are just a few of the questions we will be debating for many years to come.

Chapter 16

Transcribe This! RNA Transcription

In This Chapter

- Finding out what's in your genes
- Breaking the genetic code
- Modeling gene regulation

Cells utilize a number of types of ribonucleic acid — RNA.

Messenger RNA (mRNA), a form that is not very stable, carries information from the cell nucleus (DNA) into the cell and must migrate to the ribosomes. Messenger RNA carries the actual genetic information necessary for the synthesis of a specific protein; however, the other forms of RNA are necessary to complete the process.

Transfer RNA (tRNA) transfers amino acids to the ribosomes for protein synthesis. This is a relatively small form of ribonucleic acid, typically containing from 73 to 93 nucleotides.

The relatively large ribosomal RNA (rRNA) resides in the ribosomes and has a direct influence on the synthesis of proteins. This form of RNA has protein components. There are three types of rRNA (called 23S, 16S, and 5S), and all three must be present in each ribosome.

Finally, small nuclear RNA (snRNA) serves a number of ancillary functions. In this chapter we concentrate upon the synthesis of RNA, which is called transcription.

Protein synthesis begins with *transcription,* the process whereby DNA produces mRNA. First, a portion (a gene) of a DNA double helix opens. Nucleotides can then bind to the exposed DNA nucleotides through a process similar to replication. However, this process differs from replication in that only a portion of the

DNA opens, and the entering nucleotides contain uracil in place of thymine. One gene yields one mRNA, which, in turn, may lead to the synthesis of one or more proteins.

The enzyme RNA polymerase joins the nucleotides to produce RNA in a process that occurs within the cell nucleus. The process begins as an initiation signal toward the 5' end of RNA and goes toward a termination sequence nearer the 3' end.

RNA Polymerase Requirements

Three requirements are needed for RNA polymerase to operate. It requires activated precursors of each of the four ribonucleoside triphosphates (ATP, CTP, GTP, and UTP) from which to produce the new RNA. (See Figure 16-1.) A divalent metal ion, either magnesium or manganese, is necessary. Finally, a template must be present. Single-stranded DNA will work; however, the preferred template is double-stranded DNA. However, the DNA strands must open (separate) in order to allow the RNA polymerase access.

There are many similarities between replication and transcription. In both processes, the direction of synthesis is 5' → 3'. Elongation occurs as the 3'-OH group of the chain attacks the innermost phosphate of the entering nucleoside triphosphate. This is called a *nucleophilic attack.* The hydrolysis of pyrophosphate provides the impetus to drive the process forward. However, there are differences. Unlike its DNA counterpart, RNA polymerase is not capable of "reviewing" its work and then eliminating a mismatched nucleotide. RNA polymerase does not require a primer.

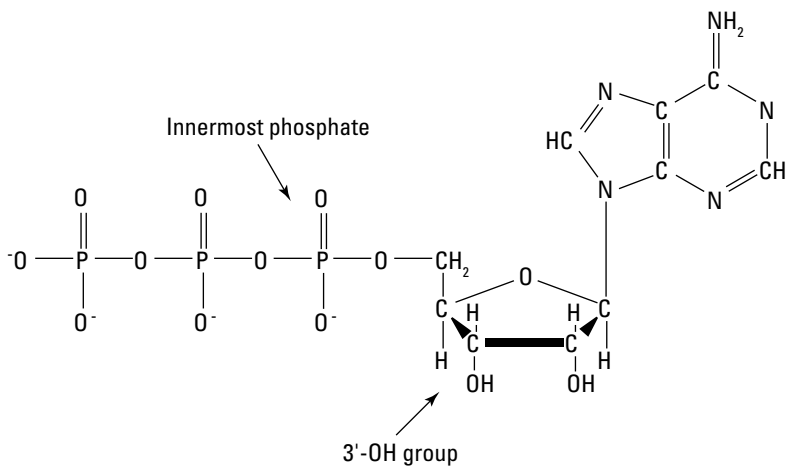


Figure 16-1: Structure of ATP.

In simple organisms, such as *E. coli*, one type of RNA polymerase synthesizes all forms of RNA. In more advanced organisms, like human beings, there are different types of RNA polymerase. Usually at least three different types are present in mammalian cells.

Making RNA: The Basics

The region of a DNA molecule that codes for a protein is a *structural gene.* Other regions are present to regulate the activity of this gene. (We examine these regulatory regions later in this chapter.) To begin transcription, it is necessary for RNA polymerase to detect one particular gene present in a long DNA strand. Detection begins with the enzyme locating a region on the DNA strand known as a *promoter site,* which is "upstream" from the actual gene. (Upstream means on the 5' side.) RNA polymerase tightly binds to the promoter site, and once in place, transcription can begin.

In prokaryotic cells the promoter sites are centered at –10 (the Pribnow box) and in the –35 region. The Pribnow box has the consensus sequence TATAAT centered at –10. The other site has the consensus sequence TTGACA. (Not all organisms have the same consensus sequence.) In eukaryotic cells a promoter is centered at about –25 (the TATA box or Hogness box), and sometimes centered near –75 (the CAAT box). The consensus sequence in the Hogness box is TATAAA. The CAAT box has the sequence GGXCAATCT. In addition, eukaryotic genes may have enhancer sequences up to several thousand bases away from the start site and on either side (Figure 16-2).

The position of sequences along the DNA chain begins at the beginning of a gene. This position is 0. The first nucleotide of the gene is +1. Counting upstream (towards the 5' terminus) is negative. Thus, ten nucleotides before the beginning of the gene would be –10.

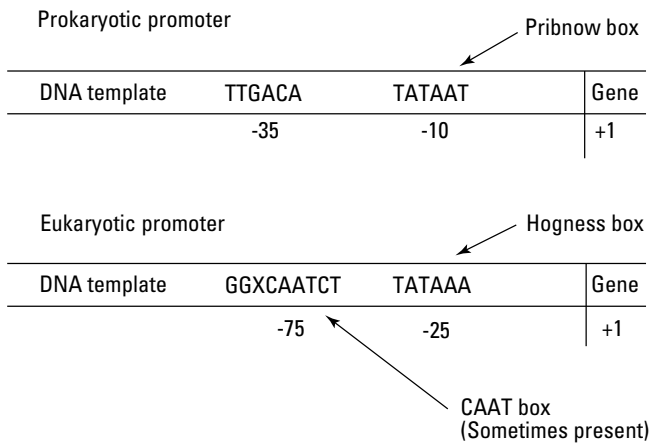


Figure 16-2: Prokaryotic and eukaryotic promoter sites.

Transcription proceeds as an RNA polymerase moves along the DNA strand. Eventually, the enzyme encounters a termination signal. In prokaryotic cells, there are two termination signals. The first is a *base-paired hairpin,* which consists of a self-complementary sequence rich in C and G followed by a sequence of several instances of U. After the sequence forms, the new RNA detaches from the template. The other method uses a *rho protein.*

The termination in eukaryotic cells is not very well understood. In eukaryotic cell, mRNA undergoes further modification after transcription. A "cap" is attached to the 5' end of the RNA, and a poly(A) tail goes onto the other end. These modifications increase the lifetime of mRNA.

The stages in RNA synthesis are *initiation*, *elongation*, and *termination*. To accomplish these tasks, RNA polymerase must perform a series of functions. The enzyme must travel along a DNA strand until it encounters a promoter site. As it "sticks" to the promoter site it unwinds a short segment of the DNA double helix and separates the strands to reach the template. Then the appropriate ribonucleoside triphosphate enters, and hydrolysis of the phosphate occurs in order to supply the needed energy. Each ribonucleoside triphosphate is brought in as the RNA polymerase moves along the DNA strand. (The DNA unwinds as the enzyme passes, and rewinds after the enzymes has passed.) This continues until the RNA polymerase finds a termination signal. The enzyme also must interact with transcription factors or trans-acting factors — proteins that act as activators or repressors — to regulate the rate of transcription initiation.

The best understood operation of RNA polymerase comes from studies of the prokaryotic cells of *E. coli.* Eukaryotic cells behave in a similar, though more complicated, manner. One major difference between the two is that in prokaryotic cells, transcription and translation (protein synthesis) may occur almost simultaneously, whereas in eukaryotic cells there is a gap between the two processes while the mRNA moves from the nucleus to the ribosome. The other major difference is that RNA in eukaryotic cells almost always requires processing after synthesis. Prokaryotic RNA is usually ready immediately after synthesis. Processing includes adding a cap, adding a poly(A) tail, and — in nearly all cases — splicing to remove introns.

Prokaryotic cells

RNA polymerase in *E. coli* contains four subunits that combine to form a holoenzyme designated $\alpha_2\beta\beta'\sigma$. The purpose of the σ subunit is to help find the promoter and to help initiate RNA synthesis. Once synthesis begins, this unit leaves the remainder, the core enzyme. The catalytic site in the core enzyme contains two divalent metal ions, one of which stays with the core and one that enters with the ribonucleoside triphosphate and leaves with

the cleaved pyrophosphate. Three aspartate residues aid in the binding of the metal ions. Although DNA polymerase and RNA polymerase have very different overall structures, their active sites are similar.

In the absence of the σ subunit, RNA polymerase would bind tightly to DNA at any point. When this unit is present, binding at other than a promoter site is significantly lower. Due to its reduced affinity, the holoenzyme can slide along the DNA strand until a σ subunit detects a promoter site. It binds to this site more strongly than to other positions on the DNA strand. The efficiency of this binding is one form of regulation. A number of σ subunits are present, each designed to recognize a different promoter site.

It's almost like tying knots in an anchor rope. A diver could swim upward holding on to the rope, but a knot signals a spot to stop and decompress.

Once the RNA polymerase arrives at a promoter site, it becomes necessary to unwind a 17 base-pair segment of the double helix and to unpair the bases. This unwinding converts a closed promoter complex to an open promoter complex. RNA polymerase is now ready to begin the RNA chain by incorporating the first nucleotide triphosphate. (Unlike DNA replication, no primer is necessary.) This first nucleotide triphosphate is usually a pppG or a pppA, which remains throughout transcription. This tap is at the 5' end of the new RNA molecule, and growth begins when a new nucleotide links to the 3' position (Figure 16-3).

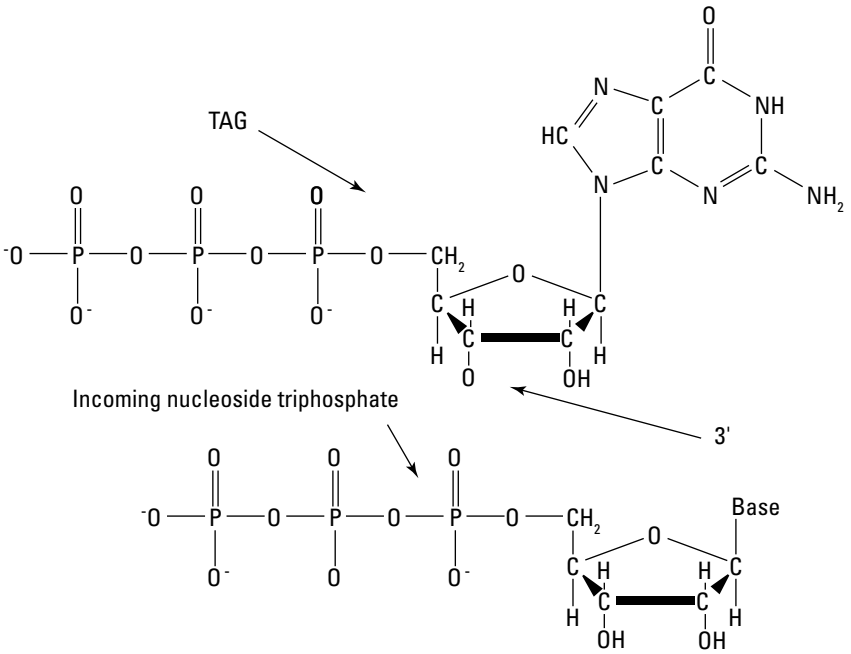


Figure 16-3: Linking of the second nucleotide to the tag, using pppG as an example (top), and linked nucleotides at the beginning of the chain (bottom).

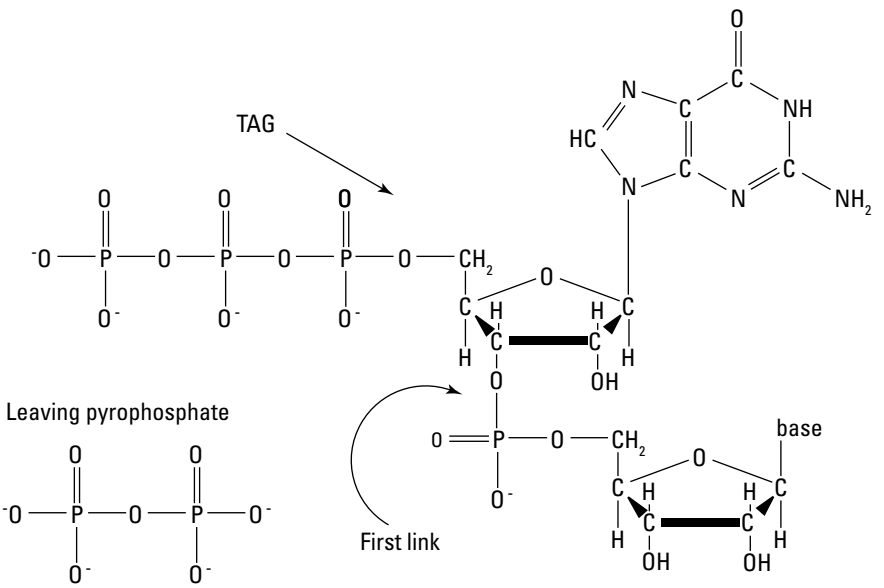


Figure 16-3: continued

Once the first two nucleotides link (through the formation of the linking phosphate diester) the σ subunit leaves. This allows the core enzyme to bind more tightly to the substrate. A transcription bubble now forms that contains the RNA polymerase, the unwound portion of the DNA, and the rapidly forming nascent RNA. Initially a short segment of the new RNA forms a hybrid helix with the DNA. This segment normally consists of about eight base pairs or one turn of the double helix. The growth rate is on the order of 50 nucleotides per second. (Compare this to DNA replication, which proceeds at about 800 nucleotides per second.) It is important to note that RNA polymerase does not "proofread" the new RNA. Thus, errors creep in at a higher rate than in replication. However, because the products do not pass to the next generation, there is no mutation or lasting effect. In any case, the next RNA stand to form lacks this defect and behaves correctly. One bad RNA in several hundred or more copies of the same gene is likely to have a minimal influence on the cell.

Elongation proceeds until the RNA polymerase encounters a termination signal, initiating a series of actions. At this point, formation of new phosphate diesters ceases, the RNA-DNA hybrid separates, the portion of the DNA chain that is still open rewinds, and the RNA polymerase separates from the DNA. There are different termination signals. One simple one is a palindromic (reading the same forward or backwards) GC-rich region followed by an AT rich region. The palindromic region is self-complementry, and these bases hydrogen bond to form a hairpin loop. The AT-rich region results in a number of U_{RNA}-A_{DNA} pairs, which have the weakest hydrogen bond interactions of all types of pairs. The formation of this hairpin and the AT region destabilizes, and the RNA-DNA hybrid and the nascent RNA begin to leave. See Figure 16-4.

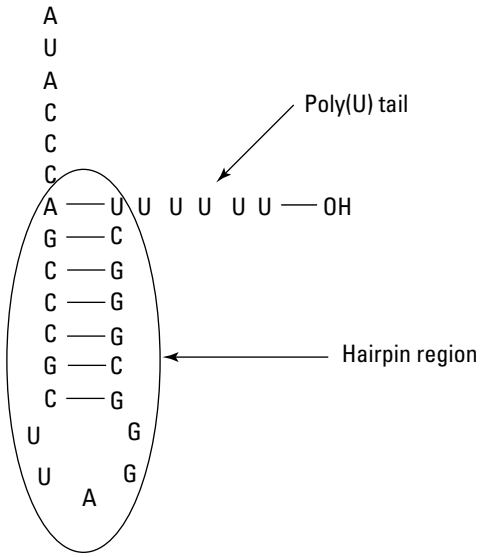


Figure 16-4: The hairpin and subsequent portion of the RNA.

Not all termination signals contain a hairpin and a U-rich segment. In at least some cases, RNA polymerase needs help. Evidence for this came from the observation that in vitro RNA chains were often longer than in vivo chains for the same RNA. Clearly, the in vitro RNA polymerase was unable to terminate elongation. The missing aid was a protein known as the *rho factor* (ρ). This protein wraps about the nascent RNA soon after the RNA exits the transcription bubble. In the presence of RNA, the ρ protein hydrolyzes ATP, which supplies energy. The protein first attaches to an RNA segment that is poor in guanine and rich in cytosine. Rho moves along the nascent RNA until it encounters the transcription bubble. At this point, it breaks the RNA-DNA hybrid and separates the nascent RNA. Other proteins serve a similar function as the rho factor.

In prokaryotic cells, mRNA is either ready or nearly ready to function immediately after release from the transcription (translation may begin before transcription terminates). However, both tRNA and rRNA require cleavage and other modifications of the nascent RNA chain. Various nucleases cleave the RNA in a very precise manner. It is possible to get more than one gene from a long nascent RNA strand. Processing may require the connection of a number of nucleotides — for example, all tRNA molecules need a CCA tail to function correctly. In some cases, there may be modification of the bases or ribose units.

Eukaryotic cells

Unlike prokaryotic cells, transcription and translation occur in different regions of the cell, leading to greater control of gene expression. Another difference is that eukaryotic cells extensively process mRNA in addition to rRNA

and tRNA. After RNA polymerase action, mRNA acquires a cap and a poly(A) tail. Nearly all mRNA molecules are spliced. Splicing involves removal of introns with the remaining exons being connected. Ninety percent of the nascent RNA may be introns.

Eukaryotic cells typically contain three types of RNA polymerase. Type I RNA polymerase (in the nucleolus) produces most forms of rRNA. Type II (in the nucleoplasm) produces mRNA and snRNA. And type III (in the nucleoplasm) produces tRNA and small rRNA molecules. (Actually, these polymerases only produce the pre-RNA forms of these molecules.)

Each of the three polymerases has a distinct type of promoter. These promoters may be in the same upstream sites as in prokaryotic cells, in downstream sites, or within the genes themselves. In addition to promoters, there may be enhancers. Enhancers, though not promoters, increase the effectiveness of a promoter. Enhancers for a single promoter may occur in different positions on the DNA chain and are important for gene regulation. Both promoters and most enhancers are on the same side of the DNA chain as the gene they regulate; for this reason, they are cis-acting elements. The promoters, as discussed earlier, are typically a TATA box (usually between –30 and –100), the CAAT box, and the GC box (both are usually between –40 and –150). Enhancers may appear upstream, downstream, or within the gene about to undergo transcription. Enhancers that are present on the opposite DNA chain are trans-acting factors, known as transcription factors, on the other DNA chain.

The typical series of events is that the transcription factor TFIID binds to the TATA box (TF stands for *transcription factor,* and the II means *RNA polymerase II*). Binding is the result of a small component of TFIID known as TBP (TATA-box-binding protein), which has an extremely high affinity for the TATA-box. When TBP binds to the DNA, substantial changes occur in DNA, including some degree of unwinding.

Other components utilized in transcription later attach to the TBP. These are, in order: TFIIA, TFIIB, TFIIF, RNA polymerase II, and finally TFIIE. This final group is the basal transcription complex. This example illustrates only one of numerous transcription factor initiations.

In eukaryotic cells, nearly all, if not all, products of transcription (precursors) undergo further processing before they reach their final active form. In general, tRNA precursors need to have the 5' leader removed, splicing to remove any and all introns, replacement of the poly(U) tail with a CCA sequence, and possible modification of some of the bases. Each of these processes requires one or more enzymes.

The precursors to the various forms of mRNA normally require the most modification. These precursors need, amongst other things, a 5' cap and a 3' poly(A) tail. The caps are cap 0, cap 1, and cap 2 — the numbers refer to the number of methylated ribose sugars (Figure 16-5). Caps are not present on tRNA, snRNA, or rRNA.

Most mRNA has a poly(A) tail not encoded by DNA. Usually, addition of this tail is preceded by cleavage of an intron portion of the mRNA precursor. The series AAUAAA signals where the cleavage will occur. This series is only part of the signal — the other part is uncertain. After cleavage, a poly(A) polymerase adds about 250 adenylate residues to the 3' end. The exact purpose of the tail is uncertain. It appears to enhance translation and increase the lifetime of the mRNA molecule.

In some cases, it is necessary to edit some mRNA precursors. *Editing* refers to an alteration of the base sequence other than that caused by splicing. An example is to chemically change one base into another. An example of editing occurs in the mRNA that encodes for apolipoprotein B (apo B). The entire protein contains 4,536 residues. However, a related 2,152-residue form is also important. The longer form, synthesized in the liver, is useful in the transport of lipids within the liver. The smaller form, synthesized in the small intestine, interacts with dietary fats. The same mRNA is responsible for both forms of the protein. In the small intestine, a deaminase acts on a specific cytosine and converts it to a uracil, which changes a CAA codon (Gln) to a UAA codon (stop) — which truncates the protein chain to yield the smaller form.

Splicing is a very common form of modification of all forms of RNA. *Splicing* involves the removal of introns and the joining of the exons to yield the final RNA molecule. Splicing must be very precise, as a miss by one base alters the entire sequence of codons present.

A number of different introns need to be removed. In eukaryotic cells, the intron begins with a GU and ends with an AG. Further refinement is present in vertebrates, where GU is the end of the sequence AGGUAAGU. A variety of AG sequences are found in higher eukaryotic cells. In general, one end of the intron loops about and connects to a point (the branch point) on the intron chain. Joining of the exons present then proceeds.

Spliceosomes are important in the splicing of mRNA precursors. These assemblages contain the mRNA precursors, several snRNAs, and proteins known as splicing factors. A group of snRNAs labeled U1, U2, U4, U5, and U6 are important. U1 binds to the 5' end of the splice site and then to the 3' end. U2 binds to the branch point, U4 blocks U6 until the appropriate moment, U5 binds to the 5' splice site, and U6 catalyzes the splicing. There are alternate splicing procedures.

It should be noted that alternate splicing can lead to production of different proteins from the same RNA.

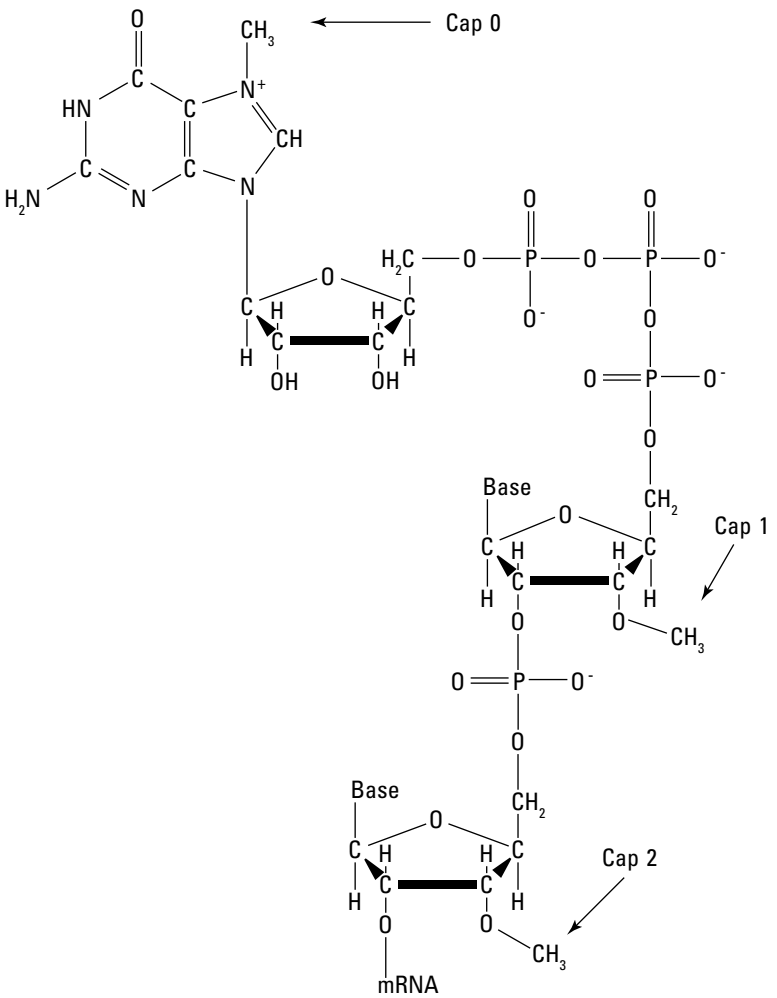


Figure 16-5: The general structure of a mRNA cap.

To Heck with Da Vinci: The Genetic Code

Just as DNA serves as the template for the generation of RNA, mRNA serves as the template for the generation of protein. In order to synthesize the appropriate protein, there needs to be a species that interacts with this template to assure the incorporation of the correct amino acid. The interaction species is tRNA. This relatively small form of RNA has two important regions: a template recognition site and the appropriate amino acid. The template recognition site is an anticodon, which corresponds to a codon on the mRNA. Attachment of the amino acid to the tRNA is by the action of an aminoacyl-tRNA synthetase. Each of the 20 amino acids has at least one

specific synthetase. This enzyme attaches the specific amino acid to the 3' terminal adenosine of the tRNA (Figure 16-6).

Codons

The genetic code contains the information necessary for the synthesis of proteins and consists of a set of three-letter words made from an alphabet containing four letters. Each three-letter word is a *codon.* This vocabulary is universal as it applies to all known living organisms.

The four letters are as follows:

- A, for adenine
- C, for cytosine

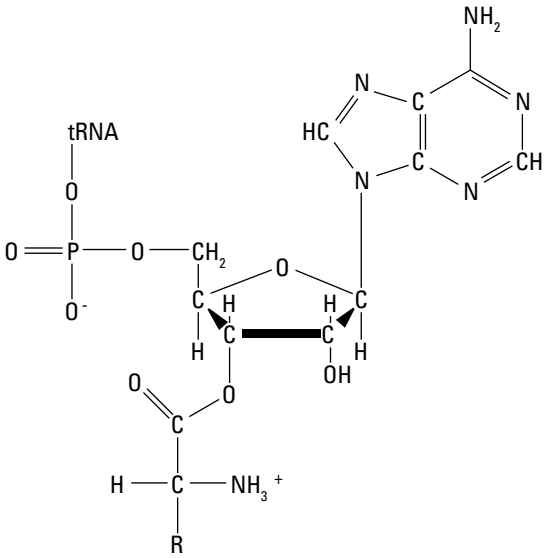


Figure 16-6: The attachment of an amino acid to the terminal adenosine.

- G, for guanine
- U, for uracil

The four letters give a total dictionary containing 64 words. Sixty-one of these words code for specific amino acids, and the remaining three words code for no amino acid. The codons coding for no amino acid are the "stop" signals.

Because there are only 20 amino acids to code for, the presence of 61 codons means that some amino acids can come from more than one codon.

Table 16-1 lists the genetic code.

Table 16-1 The Standard Genetic Code

Codon	*Amino Acid*	*Codon*	*Amino Acid*	*Codon*	*Amino Acid*	*Codon*	*Amino Acid*
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
UUA	Leu	UCA	Ser	UAA	***Stop***	UGA	***Stop***
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUG	Leu	UCG	Ser	UAG	***Stop***	UGG	Trp
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys

Analysis of the genetic code shows that two amino acids — methionine (see Figure 16-7) and tryptophan — only have one codon each. At the other extreme, three amino acids — arginine, leucine, and serine — each have six codons. The remaining 15 amino acids have at least two codons each. Amino acids with more codons are more abundant in proteins. Examining Table 16-1 shows that most *synonyms* (codons coding for the same amino acid) are grouped together and differ by a single base, usually the last base in the codon. Other correlations are present in the table. See what others you can find. The similarity of synonyms limits potential damage due to mutations.

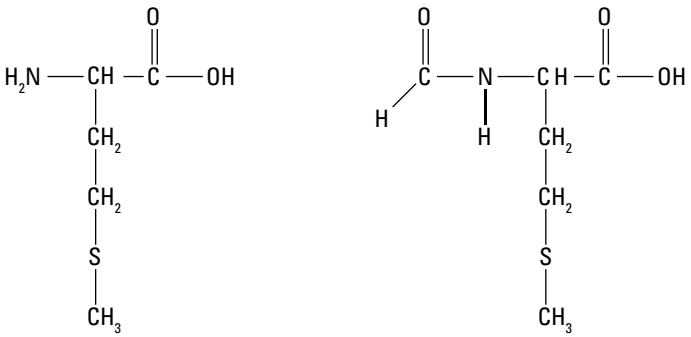


Figure 16-7: Structures of methionine and formylmethionine.

Alpha and omega

Although tRNA does not read the termination sequences, UAA, UAG, and UGA, specific proteins known as *release factors* read them. When a release factor binds to the ribosome, it triggers the release of the new protein, and release of the protein signals new synthesis to begin.

The stop signals are rather obvious on the Table 16-1, but what about the start? What signals the initiation of protein synthesis? The initiation sequence is usually AUG, the codon for methionine (Figure 16-8). In eukaryotic cells, additional factors come into play. In many bacteria, fMet (formylmethionine) is the initial amino acid (refer to Figure 16-7), which AUG usually codes for; however, GUG works sometimes.

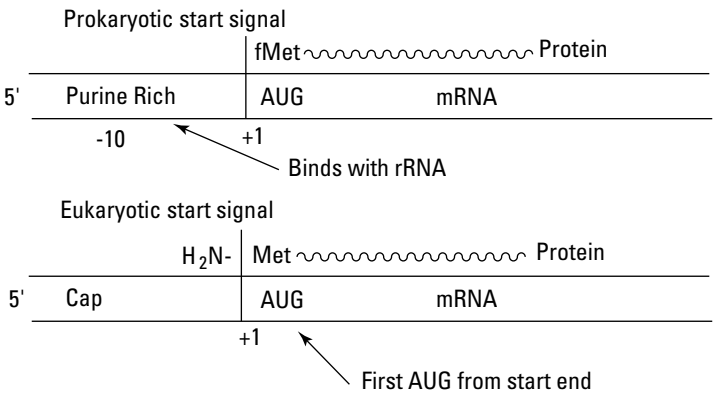


Figure 16-8: The start signals.

The genetic code is nearly universal — the codons correspond to the same amino acid in most cases. A few exceptions are known. For example, the code in mitochondrial DNA has several differences from normal DNA. In mitochondrial DNA, UGA is not a stop signal, but a codon for tryptophan.

In prokaryotic cells, coding for proteins is continuous, but this is not always true in the case of eukaryotic cells. In some mammals and birds, most genes are discontinuous. For example in the gene encoding for β-globin, there are regions that do not encode for a portion of protein. The gene contains about 1,660 base pairs — about 250 pairs on each end, plus an additional 500 pair segment code for the protein. These coding segments are *exons.* Two segments, one of about 120 base pairs and one of about 550 base pairs, do not code for protein. These non-coding regions are *introns.* The entire gene has, in sequence, a 240-pair exon, a 120-pair intron, a 500-pair exon, a 550-pair intron, and a 250-pair exon.

If an mRNA forms from a gene containing introns, it needs to undergo modification before it is of use. It is necessary to cut the intron regions from the mRNA and to splice the exon ends together to form the final mRNA molecule. In most cases, the intron portion begins with a GU and ends with a pyrimidine-rich segment ending with an AG. This combination signals the intron domain.

Models of Gene Regulation

The organism does not need to produce all the different proteins all the time. To control which proteins form at which time requires some form of *gene regulation.* When the organism requires a specific protein, it is necessary to "switch on" a certain gene — and once there is a sufficient quantity of that protein the gene must be "switched off." Control may occur either at the transcription level (gene regulation) or at the translation level.

In this section we examine processes in prokaryotic cells and then move on to the more complicated processes that take place in eukaryotic cells. The examination of the simpler mechanisms in prokaryotic cells gives insight into the processes in eukaryotic cells — the basic processes are similar.

As usual, our prokaryotic example is *E. coli.* Insight on gene regulation came when the diet of the *E. coli* was changed from glucose-rich to lactose-rich. For the cells to utilize this alternate energy source, they must generate the enzyme β-galactosidase. This enzyme is normally available at very low levels — a situation that quickly changes after replacing the glucose with lactose. One clue to the mechanism was that as the levels of β-galactosidase increased, so did the levels of galactoside permease (which transports lactose into the cell) and thiogalactoside transacetylase (which detoxifies other materials transported by galactoside permease). Thus, one change in the environment triggered multiple enzymes. This coordinated triggering of gene expression is

called an *operon*. Francois Jacob and Jacques Monad proposed the operon model to explain gene regulation.

The Jacob-Monod (operon) model

The simultaneous change in the levels of three different enzymes by one change in the environment suggested a link between the control mechanisms, and the operon model was created to account for this link. This model requires a regulator gene that affects a number of structural genes and an operator site. The operator and associated structural genes constitute the operon. The regulator gene is responsible for producing a *repressor protein*. The repressor protein binds to the operator site and prevents expression of the structural genes, as shown in Figure 16-9. The lac operon is one of the better understood operons.

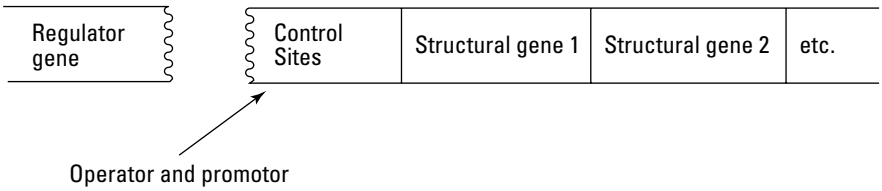


Figure 16-9: Diagram of a generic operon.

The multiple structural genes produce one large mRNA, and this single RNA strand is capable of generating a set of proteins. An mRNA that is capable of encoding for multiple proteins is *polygenic* or *polycistronic*.

The lac operon

The lac operon is the model regulatory system that, since its discovery in 1961, has provided extensive insight into how a cell regulates its genome. Figure 16-10 illustrates the lac operon.

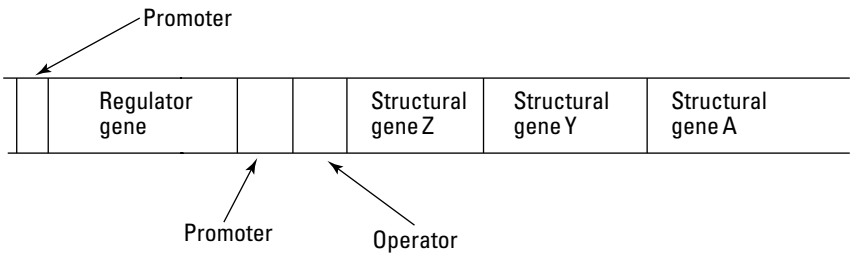


Figure 16-10: Diagram of the lac operon.

The *lac operator* is a palindromic DNA sequence with a twofold symmetry axis. The repeat is not always a perfect palindrome. (Many protein-DNA interactions involve a matching of symmetry.) The lac operator is as follows, with the center axis in bold:

TGTGTGGAATTGTGAGC**G**GATAACAATTTCACACA

ACACACCTTAACACTCG**C**CTAATGTTAAAGTGTGT

The *lac repressor* is a dimeric protein that can join to form a tetramer. In the absence of lactose, the repressor tightly binds to the operator. The presence of the repressor prevents RNA polymerase from unwinding the DNA strand to initiate transcription.

The presence of lactose is not the direct trigger of the lac operon; the trigger is allolactose. Both lactose and allolactose are disaccharides composed of galactose and glucose (Figure 16-11). In lactose there is an α-1,4 linkage, whereas in allolactose the linkage is an α-1,6. Allolactose results when the few molecules of β-galactosidase that are normally present in the cell first encounter lactose. This disaccharide along with a few similar molecules is an inducer of the lac operon. The inducer binds to the repressor and reduces the affinity of the latter to operator on the DNA. With its affinity reduced, the repressor detaches from the operator, and the DNA segment is now open for business.

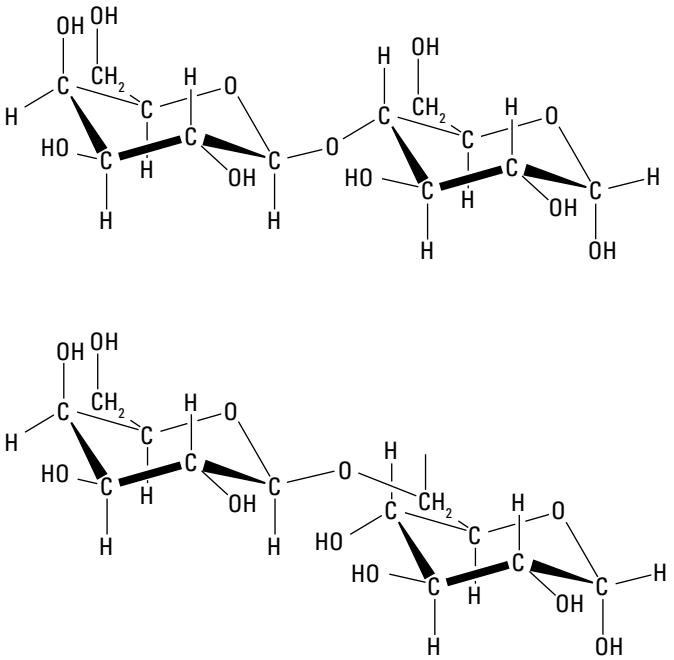


Figure 16-11: Structures of lactose and allolactose.

When transcription begins, all three structural genes become active, and the cell begins producing β-galactosidase, galactoside permease, and thiogalactoside transacetylase. This continues until the lactose and hence the allolactose concentration falls so that the repressor proteins are available to reattach to the DNA.

Other prokaryotic regulators

The *pur repressor* affects the genes responsible for the biosynthesis of purines and, to a lesser extent, pyrimidines. This protein is similar in structure to the lac repressor; however, the pur repressor only binds to the operator after another molecule binds to the repressor. Therefore, while the binding of another molecule releases the lac repressor, the binding of another molecule causes the pur repressor to bind. The other molecule has an opposite affect. In the case of the pur repressor, the other molecule is a *corepressor.*

There are also regulators that stimulate transcription instead of repressing it. The catabolite activator protein (CAP) is one example. This protein interacts with the promoter and, along with two cAMP molecules, interacts with RNA polymerase. This interaction leads to stimulating the initiation of transcription of certain genes.

Regulation of eucaryotic genes

Although there are similarities, the regulation of genes in eukaryotic cells is more complex than in prokaryotic cells. One reason for this is that the typical eukaryotic genome is much larger than the typical prokaryotic genome. Another source of complexity is that many eukaryotic cells are part of a larger organism and do not serve the same purpose as other cells do within the same organism. For example, although some of the proteins are the same, a liver cell must produce a different overall set of proteins than a heart cell does.

Histones

Eukaryotic DNA has a group of proteins associated with it. These small, basic proteins are called *histones.* They are basic because approximately 25 percent of the amino acid residues present are either arginine or lysine. These are tightly bound to the DNA and total approximately half of the mass of a chromosome. A complex of the cell's DNA and associated protein is a *chromatin,* and there are five important histones present in chromatin: H1 — and four that associate with each other: H2A, H2B, H3, and H4.

A chromatin apparently consists of repeat units consisting of two copies each of H2A, H2B, H3, and H4, with a strand of DNA consisting of about 200 base pairs tightly wrapped around this histone octomer. Each of these repeating units is a nucleosome. The wrapping of the DNA to form a nucleosome yields a significant compaction of the DNA. Research indicates that about 145 of

the 200 base pairs are actually associated with the histone octomer, and the remaining base pairs are linker DNA that link one histone octomer to the next. Histone H1 usually binds to linker DNA.

The eight histones in a histone octomer are arranged into a tetramer with the composition $(H3)_2(H4)_2$ and two dimers each with the composition (H2A)(H2B). All the histone proteins have long tails rich in arginine and lysine residues that extend out of the core. Modification of these tails is important for gene regulation.

The structure of chromatin is a factor in eukaryotic gene regulation. For a gene to be available for transcription, the tightly packed chromatin structure must open. In addition, the structure regulates access to regulatory sites on DNA. Enhancers disturb this structure, explaining why enhancers can have an effect on the expression of a gene even though the enhancer site may be thousands of base pairs away from the gene. Certain enhancers only occur in specific types of cells. Thus, the genes they enhance are only expressed in these cells. For example, the gene to produce insulin is expressed only in pancreatic cells.

A modification of DNA can also inhibit gene expression. Approximately 70 percent of the 5'-CpG-3' sequences in mammals have the cytosine methylated. The distribution of the methylated cytosines (Figure 16-12) varies with cell type. Regions in chromatins necessary for gene expression in that cell are hypomethylated (have fewer methylated cytosines), relative to similar regions in cells where no expression of the gene occurs. The presence of the methyl group interferes with the binding of enhancers and promoters.

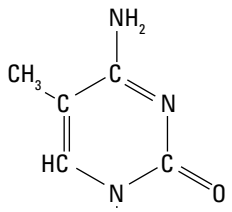


Figure 16-12: Structure of methylated cytosine.

Mediating transcription

Eukaryotic cells require a variety of transcription factors to initiate transcription — no factor can carry out the entire process on its own. This is called *combinatorial control,* and it is necessary in organisms with multiple cell types and helpful in other eukaryotic cells.

A number of nonpolar molecules, such as the steroid hormones, can easily pass through the hydrophobic cell membrane and bind to receptor proteins. They are very specific. Estrogen (Figure 16-13) is one example of a steroid hormone. Such molecules are known as *ligands.*

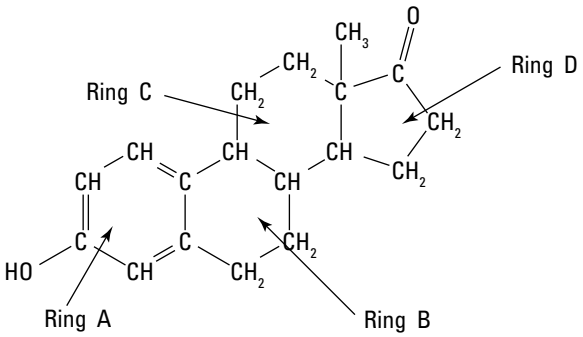


Figure 16-13: Structure of estrogen.

The ligand binds to a specific site — called, helpfully, the *ligand-binding site* — which is present near the end of a receptor protein. This portion of the protein contains many nonpolar residues, which have an affinity for hydrophobic molecules. Receptor proteins that bind hormones are called *nuclear hormone receptors.* There is a DNA binding site near the center of the protein that contains eight cysteine residues, which are necessary to bind zinc ions, four residues for each. The presence of the zinc ions stabilize structure and led to the name *zinc finger domains.* (There are other cysteine residues and zinc ions nearby.) The binding of a molecule to the ligand-binding site causes a significant structural rearrangement of the protein. This situation would seem to be similar to the lac repressor in prokaryotic cells; however, experiments indicate that there is no significant alteration in binding affinity

The next part of the puzzle involves a number of small proteins known as *coactivators.* Near the center of each of these are three regions with the pattern Leu-X-X-Leu-Leu. Each of these regions generates a short hydrophobic α-helix. These three helices bind to a hydrophobic region on the ligand-binding region. The presence of the ligand appears to enhance the binding of a coactivator. (A receptor protein may act as a repressor, especially in the presence of a corepressor.)

Just what are the roles of coactivators and corepressors? Their effectiveness appears to be linked to their ability to covalently bond to the tails of the histones. Histone acetyltransferases (HATs) catalyze this modification of the histone tails (a process that is reversed by histone deacetylase enzymes — see Figure 16-14). This process changes a very polar (positively charged lysine) to a much less polar (neutral) amide, resulting in a significant reduction in the affinity of the tail to the associated DNA. To a lesser degree, it reduces the affinity of the entire histone to the associated DNA. The reduction in the affinity allows access of a portion of the DNA to transcription.

The acetyated lysine residues also affect the acetyllysine-binding domain (the *bromodomain*) present in many of the eukaryotic transcription regulatory proteins.

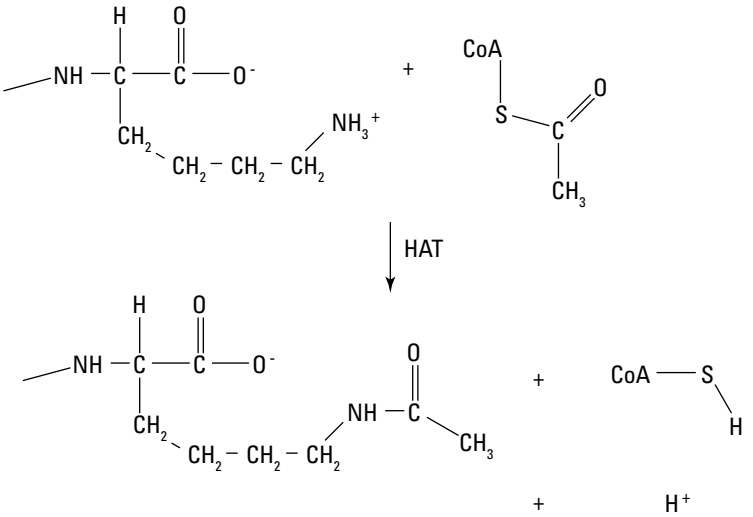


Figure 16-14: Reaction catalyzed by histone acetyltransferases (HATs).

There are two important bromodomain-containing proteins: One of these is a large complex of ten proteins that binds to the TATA-box-binding protein that is responsible for the transcription of many genes. The other proteins containing bromodomains are part of large complexes known as *chromatin-remodeling engines.* As the name implies, these proteins alter the structure of the chromatins, which changes the behavior of the chromatin.

All these factors alter the availability of portions of the DNA structure to transcription. Once the DNA becomes open, the procedures discussed earlier in the chapter come into play.

Chapter 17

Translation: Protein Synthesis

In This Chapter

- Talking about translation
- Looking at protein synthesis
- Examining eukaryotic cells
- Discussing the Human Genome Project

You are no doubt familiar with the process of translation — converting text from one language into another. The process of translation in biochemistry does exactly the same thing.

Hopefully Not Lost in Translation

Translation is the process where the four-letter alphabet of the nucleic acids becomes the twenty-letter alphabet of proteins. In doing so, genetic information is passed on. Translation occurs in the cell's ribosomes, which contain ribosomal RNA (rRNA). The information necessary for translation travels from the cell nucleus to the ribosomes via messenger RNA (mRNA). The messenger RNA binds to the smaller ribosomal body, and transfer RNA (tRNA) brings amino acids to it.

Why translation is necessary

The purpose of translation is to put together specific amino acids in a specific order to produce a specific protein. Messenger RNA provides the template or blueprint for this process. To utilize this template, something must bring the amino acids to the mRNA, and that thing is transfer RNA (tRNA). Transfer RNA has two important sites. One site is for the attachment of a specific amino acid. For example, only one specific type of tRNA will transfer the amino acid methionine. The other site is the recognition site, which contains an

anticodon. An *anticodon* is a sequence of three bases that match a codon on the mRNA. A codon sequence of AUG on the mRNA matches the UAC anticodon on a tRNA. All of this takes place in the ribosome, home of rRNA.

Home, home in the ribosome

The *ribosome* is the factory that produces proteins. Thousands of ribosomes are present in even the simplest of cells. They are complex units composed of RNA and protein. It is possible to dissociate a prokaryotic ribosome into two units. One unit is the 50S, or large unit, and the other is the 30S, or small unit. The large unit contains 34 different proteins, labeled L1 through L34, and two RNA molecules, labeled 23S and 5S. The small unit contains 21 different proteins, labeled S1 through S21, and an RNA molecule labeled 16S. The RNA molecules in the ribosomes act as transfer RNA in translation.

A prokaryotic ribosome contains three rRNA molecules (23S, 16S, and 5S), one copy of proteins S1 through S21, two copies of L7 and L12, and one each of the other L1-L34 proteins. L7 and L12 are identical except that L7 has an acetylated amino terminus. S20 and L26 are identical. Mixing the constituents in vitro leads to the two subunits reconstituting themselves. A version of the structure of the 16S form of ribosomal RNA appears in Figure 17-1.

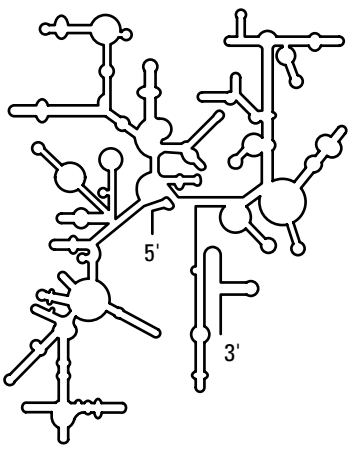


Figure 17-1: Simplified schematic of the structure of the 16S form of ribosomal RNA.

The Translation Team

A number of players, along with the rRNA, must interact in order to form a protein molecule. In addition, the structure of the ribosome is important to controlling protein synthesis. Both the rRNA and protein molecules control this structure. One possibly helpful analogy is the game of football.

The team captain: rRNA

RNA makes up approximately two-thirds of the mass of a ribosome. The three mRNA units play a key role in the shape and function of the ribosome (the proteins apparently fine-tune the shape and structure of the ribosome). The three mRNA form from the cleaving and processing of transcribed 30S RNA. A significant portion of each of the rRNA molecules have numerous duplex regions (short stretches of base-paired RNA).

The 30S and 50S rRNA subunits combine to form a 70S ribosome, which holds an mRNA in place during translation. There are also three sites for various tRNA molecules: the E, P, and A sites. The E site is the exit site. A tRNA occupies this position after delivery of its amino acid and just before exiting the ribosome. The P site is the peptidyl site, which holds the tRNA containing either the initial amino acid or the C-terminal amino acid of a protein chain. Finally, the A site is the aminoacyl site, which holds the tRNA attached to the next amino acid in sequence. When the 30S and 50S subunits join, they create A and E sites at the interface of the subunits. The P site of the 50S unit is the opening of a tunnel through which the growing protein chain passes out of the ribosome.

Here's the snap: mRNA

The base sequence of the mRNA is read in the 5' $\rightarrow$ 3' direction, and transcription occurs in this same direction. (Prokaryotic cells sometimes take advantage of this by beginning translation before transcription is over. This situation cannot occur in eukaryotic cells because the E sites of transcription and translation are physically separated.) The mRNA resulting from transcription gains a cap and a poly(A) tail before it ventures out of the nucleus on its trip to the ribosome.

Translation does not begin at the 5' terminus of the mRNA molecule. Just as there is a "stop" signal to terminate translation, there is a "start" signal. The 5' terminus base-pairs with the 3' terminus of the 16S rRNA. This region is normally about 30 nucleotides in length (a portion of this region, called the Shine-Dalgarno sequence, is purine-rich).

Shortly after this sequence is the start signal. In most cases, the start signal is AUG (methionine), though there are instances where the signal is GUG (valine). In *E. coli,* the first amino acid is formylmethionine instead of methionine. The formylmethionine is usually removed soon after translation begins. In prokaryotic cells, there may be more than one start and stop signal because many of the mRNA molecules are polygenic (polycistronic) — that is, they produce more than one protein. The structures of methionine and formylmethionine attached to tRNA are illustrated in Figure 17-2.

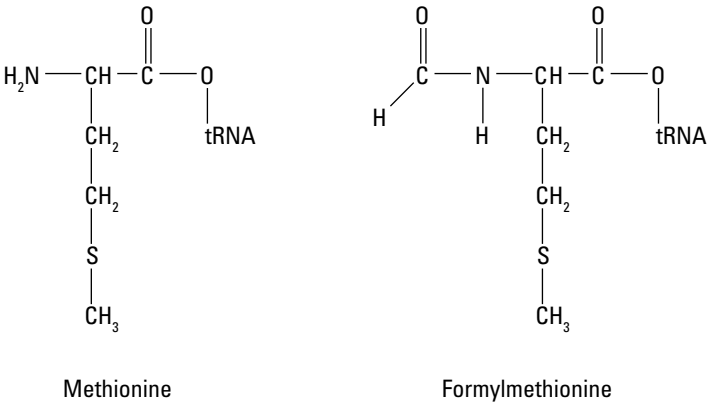


Figure 17-2: The structures of methionine and formylmethionine attached to tRNA.

Carrying the ball: tRNA

Several features are common to all forms of tRNA. Each form of tRNA is a single strand containing between 73 and 93 nucleotides. There are between seven and fifteen unusual bases (not one of the usual four, A, C, G, or U) in each molecule. Approximately half of the nucleotides present are base-paired. The activated amino acid is attached to the hydroxyl group at the 3'-end of the chain. The hydroxyl group is on the adenosine residue of a CCA segment. The other end, the 5'-end, is phosphorylated. The phosphorylation usually is a pG. The anticodon is contained in a loop near the center of the molecule.

Many of the unusual bases are methylated or dimethylated forms of A, C, G, or U, which are usually the result of post-transcription modification of the molecule. The presence of the methyl groups interferes with the formation of some base pairs, which prevents certain additional interactions. Methyl groups are nonpolar, so their presence makes regions of the tRNA hydrophobic, which affects their interaction with ribosomal proteins and syntheses. The unusual bases include dihydrouridine, dimethylguanosine, inosine, methylguanosine, methylinosine, pseudouridine, and ribothymidine. Inosine, shown in Figure 17-3, is part of the anticodon. Many of these are in or near the bends in structure of tRNA.

There are five regions, shown in Figure 17-4, that are not base-paired. (Note that the structure of tRNA shown in Figure 17-4 is not the actual three-dimensional structure of tRNA.) Starting at the 5'-end, the unpaired regions are, in order, the DHU loop, the anticodon loop, the extra arm, the TψC loop, and the 3-CCA terminus. (The name of the DHU loop derives from the presence of several dihydrouracil residues. The anticodon loop contains the segment that recognizes the codon on the mRNA, and the extra arm contains a variable number of

residues. The TψC loop derives its name from the presence of the sequence thymine-pseudouracil-cytosine.) These loops make each tRNA different, even though the overall structure is the same.

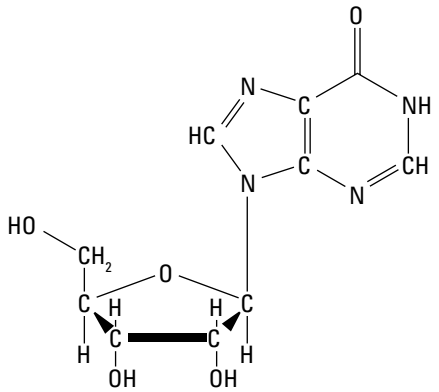


Figure 17-3: The structure of inosine.

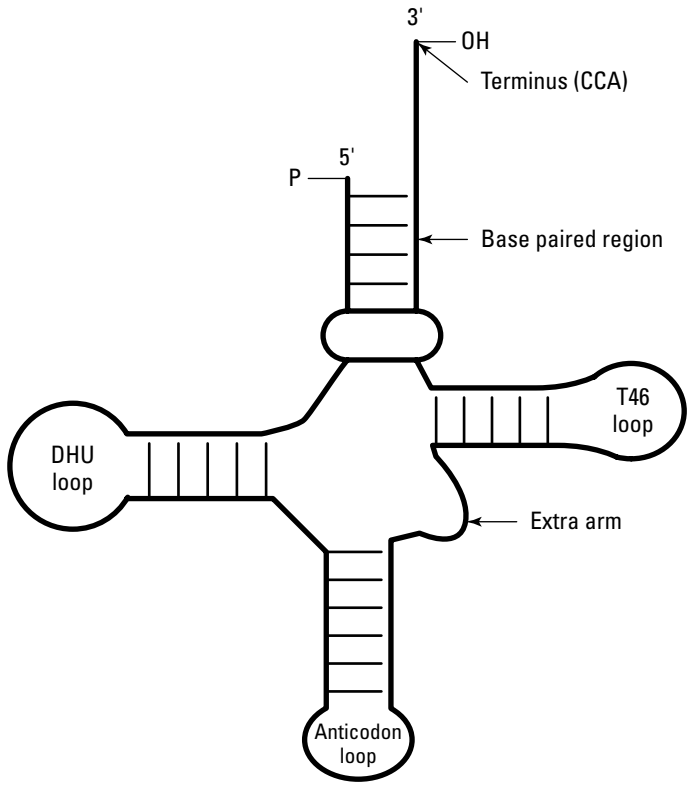


Figure 17-4: Some aspects of the structure of tRNA.

The anticodon is present in the 5' → 3' direction, and it base-pairs to a codon in the 3' → 5' direction. This matches the first base of the anticodon with the third base of the codon. (Don't forget the convention of writing base sequences in the 5' → 3' direction.)

Charging up the middle: Amino acid activation

It is imperative that the correct amino acid attaches to the tRNA because the presence of an incorrect amino acid or the absence of any amino acid would be devastating to translation. Connection of the amino acid to the tRNA activates the amino acid. Joining free amino acids is a nonspontaneous process, however, connecting the amino acid to the tRNA changes the free amino acid to a more reactive amino acid ester. The amino acid-tRNA combination is an aminoacyl-tRNA or a charged tRNA (Figure 17-5).

Specific aminoacyl-tRNA synthetases, called *activating enzymes,* catalyze the activation reaction. The process begins with an amino acid and an ATP forming an aminoacyl adenylate, which leads to the release of a pyrophosphate. Figure 17-6 shows an aminoacyl adenylate.

There is a separate aminoacyl-tRNA synthetase for each amino acid.

The two classes of aminoacyl-tRNA synthetases are denoted Class I (monomeric) and Class II (usually dimeric). Each class is responsible for ten amino acids. The CCA arm adopts different structures when interacting with members of the different classes, and ATP adopts a different conformation when interacting with members of different classes. Most Class II examples attach the amino acid as illustrated back in Figure 17-5, whereas Class I examples attach the amino acid to the alternate linking site. Some aspects of the structure of tRNA appear in the schematic structure shown in Figure 17-4.

The conversion of an aminoacyl adenylate, once formed, remains tightly bound to the synthetase until it can form an aminoacyl-tRNA.

In order to make sure that the aminoacyl-tRNA synthetase incorporates the correct amino acid, the enzyme must take advantage of specific properties of the amino acids. Examining the amino acids serine, valine, and threonine can give some insight into the selection process. These three amino acids appear in Figure 17-7, where they are drawn to emphasize similarities in the side-chain. (Recall that the threonine side-chain is chiral, but the others are not.) It is important to realize that there are size differences (–H for $–CH_3$) and

hydrogen bonding differences (–OH can, but $–CH_3$ cannot). The recognition site has the proper size and composition to take advantage of these specific properties. A significant species in this site is a zinc ion, which coordinates to the enzyme and the amino acid.

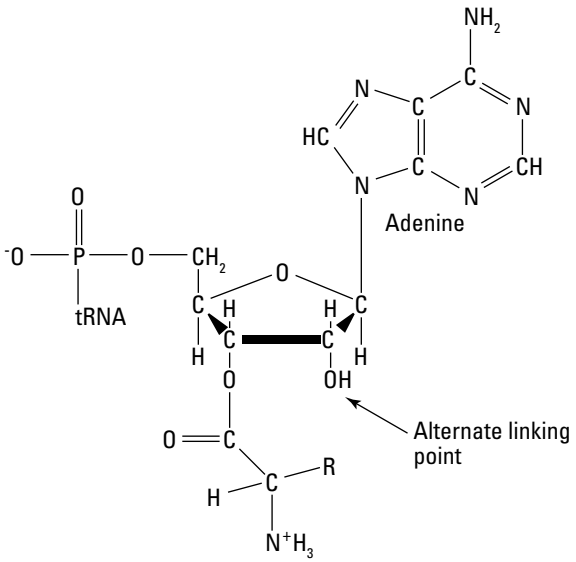


Figure 17-5: An example of an aminoacyl-tRNA.

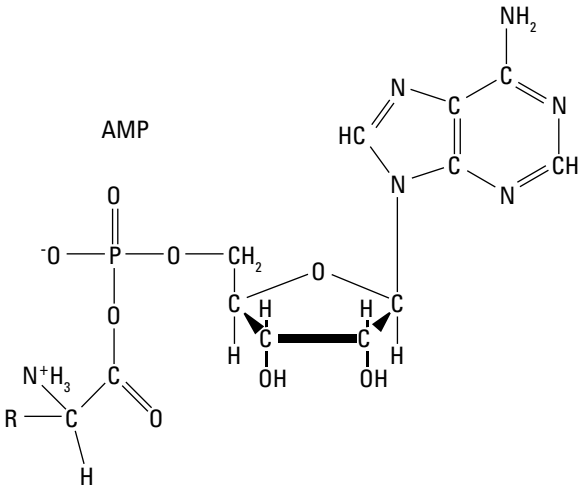


Figure 17-6: Structure of an aminoacyl adenylate.

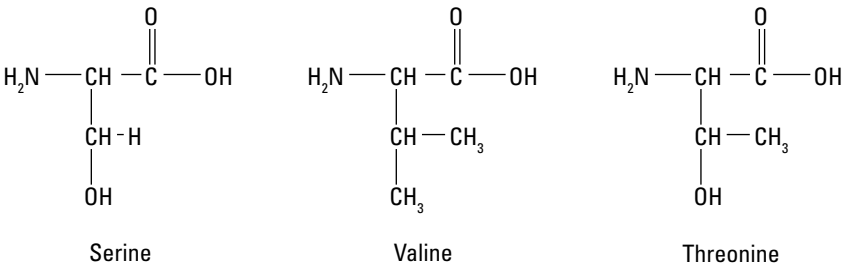


Figure 17-7: Structures of serine, valine, and threonine.

Even with these differences, serine sometimes replaces threonine. Fortunately, the enzyme includes an editing feature — the editing site is near the reaction site, but it is not the same. Similar editing occurs in other aminoacyl-tRNA synthetases. Amino acids, such as tryptophan, do not have closely similar analogues; thus, editing is far less important in these cases.

The aminoacyl-tRNA synthetases need to be able to recognize the anticodon present to make sure they interact with the appropriate tRNA, matching it to the correct amino acid. The enzymes may recognize other features of the tRNA structure. These features include the size of the extra arm and the hydrophobic character imparted by methylating some of the ribonucleotides.

Hooking Up: Protein Synthesis

The major steps in protein (polypeptide) synthesis are as follows:

- Activation
- Initiation
- Elongation
- Termination

These basics apply to all living organisms — there are no differences between human translation, fungi translation, or tulip translation. Synthesis proceeds from the amino to the carboxyl direction of the protein.

In this section we discuss these in greater detail. These steps involve tRNA, mRNA, and rRNA — along with a number of protein factors.

Activation

As mentioned earlier in this chapter, during *activation* an amino acid reacts with ATP to give aminoacyl adenylate. The aminoacyl adenylate then reacts

with a specific tRNA to give aminoacyl-tRNA plus AMP. This constitutes one of the players necessary for the translation game.

Initiation

During *initiation,* an mRNA attaches to a ribosome by interacting, through the Shine-Delgarno sequence, to the 30S rRNA subunit. Then the anticodon of the first tRNA attaches to the AUG (or GUG) codon on the mRNA. This occupies the P site of the 30S subunit. The amino acid extends into the P site of the 50S subunit. The 30S and 50S portions of the rRNA combine to produce the 70S ribosome. The combination of the two subunits allows the tRNA to interact with both parts.

In order to initiate translation it is necessary to bring the mRNA and the first tRNA to the ribosome. Three proteins, known as *initiation factors,* accomplish this task: IF1, IF2, and IF3. First, the 30S ribosome subunit, IF1 and IF3, form a complex. The two initiation factors bound to the 30S subunit interfere with a premature joining of the 30S and 50S subunit without the necessary mRNA. The remaining initiation factor, IF2, binds to GTP. The IF2-GTP combination binds to the initiator-tRNA, and the IF2-GTP-initiator-tRNA unit binds to the mRNA. Interaction of the Shine-Dalgarno sequence and the 16S rRNA manipulates the incoming group into the correct position.

Combining all these units with the 30S subunit gives the 30S initiation complex. Hydrolysis of the GTP as the 50S subunit approaches leads to expulsion of the initiation factors. With the initiation factors out of the way, the remaining moieties join to give the 70S initiation complex. (Wow, trying say that three times fast!) Once this complex forms, elongation can begin.

Elongation

During *elongation,* a second activated tRNA comes into the A site (which is adjacent to the P site) on the 30S subunit, where it binds to the appropriate codon. The activated tRNA is brought to the A site by a protein known as *elongation factor Tu* or *EF-Tu.* EF-Tu forms a complex with the activated tRNA (in the GTP form), and this complex protects the ester linkage holding the amino acid to the tRNA. In addition, the complex does not allow the activated tRNA to enter the A site if there is not a codon-anticodon match. EF-Tu interacts with all tRNAs except the initiator-tRNA. The energy needed for the EF-Tu to leave the tRNA in the ribosome comes from the hydrolysis of the GTP unit induced by the protein known as *elongation factor Ts.*

The two amino acids extend into the peptidyl transferase center of the ribosome. The amino group of the aminoacyl-tRNA from the A site is held in position to attack the ester linkage of the aminoacyl-tRNA in the P site. The catalyzed formation of the peptide bond occurs, accompanied by separation from the tRNA in the P site. The protein is now attached to the A site (30S).

With the loss of its amino acid, the tRNA no longer interacts in the same way with the ribosome. The tRNA moves to the E site of the 50S subunit as the next RNA, with its attached polypeptide, moves to the P (tunnel) site of the same subunit. The ribosome must now move over (the fancy way to say it is *translocate*) by one codon. For translocation to occur, it is necessary to employ the elongation factor G enzyme (EF-G or translocase is the protein that aids translocation). The hydrolysis of GTP to GDP supplies the energy for the move. This move places the polypeptide-tRNA into the P site of the 30S subunit. At the same time, the amino acid–stripped tRNA disengages from the mRNA and moves into the E site of the same subunit. Throughout this process, the polypeptide chain remains in the P site of the 50S subunit.

The first tRNA leaves the E site. Now the elongation cycle can begin again with the entry of another tRNA carrying the next amino acid. The process continuously cycles until a "stop" signal codon.

Termination

A "stop" signals *termination,* which results in the release of the protein, the last tRNA, and the mRNA.

Recall that the stop signals are UAA, UGA, and UAG.

Normal cells do not contain tRNAs with anticodons complementary to these codons. However, proteins known as *release factors* (RF) recognize these three codons. Release factor 1, RF1, recognizes UAA and UAG. Release factor 2, RF2, recognizes UAA and UGA. Release factor 3, RF3, is an intermediary between RF1, RF2, and the ribosome. The release factors carry a water molecule into the ribosome in place of an amino acid. The final reaction, the one that releases the newly formed protein, is the hydrolysis of the last ester linkage to a tRNA. The water brought in by the release factors is necessary for this hydrolysis.

The 70S ribosome remains together for a short time. Dissociation of the complex is mediated by a ribosome release factor and EF-G. GTP supplies the energy for this process.

The wobble hypothesis

Experimental studies have found that even pure tRNA molecules are capable of recognizing more than one codon. Biochemists developed the *wobble hypothesis* to explain this behavior, and subsequent work has firmly established this hypothesis.

The presence of the unusual base, inosine (shown back in Figure 17-3), in the anticodon loop is the key to understanding the wobble hypothesis. This base is capable of base pairing with adenine, cytosine, or uracil, allowing for some variation, or wobble, in the matching of codon to anticodon. The presence of inosine increases the number of different codons a particular tRNA can read. The first two bases in the codon pair to the corresponding bases in the anticodon. The third base is the wobble position.

Review the table of codons (Table 16-1 in Chapter 16) and see which amino acids depend only on the first two bases. Hint: Look at valine.

The base-pairing rules for the wobble hypothesis are shown in Table 17-1. The presence of an A or C as the first base allows the reading of only one codon. The presence of a G or U allows the reading of two codons, whereas an I allows the reading of three codons. Inosine is a useful base for allowing wobble; however, as Table 17-1 shows, it is only when the first anticodon base is an A or a C that there is no wobble. In general, the base in the wobble position forms weaker hydrogen bonds than normal because of the strain in the environment. The weaker hydrogen bonding aids in the loss of the tRNA after it delivers its amino acid.

Table 17-1 Base-pairing Rules for the Wobble Hypothesis

Base on Anticodon (1st Base)	*Bases Recognized on Codon (3rd Base)*
A	U
C	G
G	U, C
U	A, G
I	U, C, A

Four codons code for valine, comprising a four-codon family. If you examine three of the codons for valine: GUU, GUC, and GUA, they would all pair to the anticodon CAI instead of the anticodons CAA, CAG, and CAU. For this reason,

one CAI anticodon replaces three other anticodons. The remaining valine codon is GUG, which requires the synthesis of only two types of tRNA instead of four. Other four-codon families also work this way.

The only cases where the codons for a particular amino acid differ in the first two bases are the six-codon families, which are those of arginine, leucine, and serine. These families require three different tRNAs.

The presence of wobble reduces the number of necessary tRNAs in a cell from 61 to 31. However, cells usually have some number of tRNAs between these extremes. All the tRNAs coding for a specific amino acid require only one aminoacyl-tRNA synthetase.

Variation in Eukaryotic Cells

All cells follow the same basic pattern for translation. However, eukaryotic cells show some variations. More proteins are necessary to mediate translation, and the steps are, in general, more complicated.

Ribosomes

In eukaryotic cells, the ribosomes contain a 60S subunit and a 40S subunit, which combine to produce an 80S ribosome. The 40S subunit contains an 18S rRNA analogous to the 16S in the 30S subunit. There are three rRNA components in the 60S subunit: a 5S and a 23S, analogous to the 5S and the 23S of the prokaryotic 50S subunit, and a unique 5.8S rRNA.

Initiator tRNA

In eukaryotic cells, the initiator amino acid is methionine instead of formylmethionine. As in prokaryotic cells, a special tRNA is necessary for the first tRNA — a modification of the normal methionine-carrying tRNA.

Initiation

AUG is the only initiator codon in eukaryotic cells, and this is always the AUG nearest the 5' end of the mRNA. There is no purine-rich sequence immediately before this as in prokaryotic cells. The 40S ribosome subunit attaches to the mRNA cap and moves base by base in the 3' direction until it reaches

The Human Genome Project

The U.S. Human Genome Project was begun in 1990. It was originally scheduled to last for 15 years but because of rapid advances in the field of biotechnology it finished two years ahead of schedule in 2003. The U.S. Department of Energy and the National Institutes of Health coordinated the projects.

Goals

The Project had the following goals:

- Identify all the 20,000–25,000 genes in human DNA.
- Determine the sequences of the approximately 3 billion base pairs in human DNA.
- Store the information in databases.
- Improve data analysis tools.
- Transfer the developed technology to the private sector.
- Address the ethical, legal, and social issues associated with the project.

In addition to human DNA, researchers also studied the genetic blueprints of *E. coli*, a common bacterium found in humans as well as mice and fruit flies. The goal of transferring the technology to the private sector was included to develop the infant biotechnology industry and encourage the development of new medical applications.

Potential Benefits

Some potential benefits of the Human Genome Project include the following:

- Improved disease diagnosis
- Earlier detection of genetic predispositions to disease
- Drug design and gene therapy
- Creation of new biofuels
- More effective ways of detecting environmental pollutants
- Studying evolution through mutations in lineages
- Forensic identification of subjects through DNA analysis
- Establishing paternity
- Matching organ donors and patients
- Creation of insect- and disease-resistant crops
- Creation of biopesticides
- Increased productivity of crops and farm animals

Many of these potential benefits are showing up in our everyday life already.

Ethical, Legal, and Social Issues

One of the unique aspects of the Human Genome Project was that it was the first large scientific project that studied and addressed potential ethical, legal, and social implications that arose from the data generated from the study. Questions such as the following were addressed:

- Who should have access to personal genetic information?
- Who controls and owns genetic information?
- How reliable and useful is fetal genetic testing?
- How will genetic tests be checked for reliability and accuracy?
- Do parents have the right to test their children for adult-onset diseases?
- Do people's genes influence their behavior?
- Where is the line between medical treatment and enhancement?
- Are genetically modified foods safe for humans?

Many questions have been raised — but, as yet, few answers have resulted.

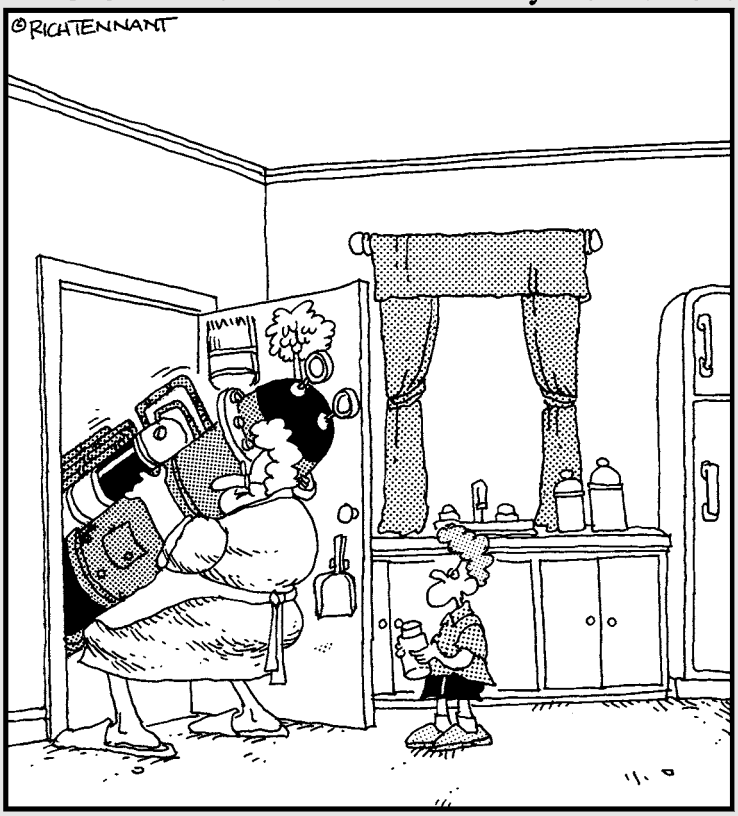
an AUG codon. The hydrolysis of ATP by helicases powers this process. Many more initiation factors are present in eukaryotic cells. A eukaryotic initiation factor has the symbol eIF instead of IF.

Elongation and termination

The EF-Tu and EF-Ts prokaryotic elongation factors have the eukaryotic counterparts EF1α and EF1$\beta\gamma$. Translocation is driven by eukaryotic EF2 with the aid of GTP. Only one release factor, eRF1, is present in eukaryotic cells instead of the two factors in prokaryotic cells. To prevent the reassembly of the two ribosome subunits, eIF3 functions like the IF3 protein in prokaryotic cells.

Part VI
The Part of Tens

The 5th Wave By Rich Tennant



"You can take that old jar for your science project, I'm sure I have some baking soda you can borrow, and let's see, where's that old particle accelerator of mine...here it is in the pantry."

In this part . . .

We wrap things up by zooming out a bit and looking at things we haven't covered yet. Here we compile two chapters' worth of short and sweet information about some of the lesser-known potential applications of biochemistry and some perhaps unexpected careers related to it.

Chapter 18

Ten Great Applications of Biochemistry

In This Chapter

- Examining tests
- Checking out genetically modified foods
- Considering cloning

In this chapter, we briefly look at some of the biochemical applications and tests that have changed our everyday lives. Although these are just a few of the hundreds of examples we could have chosen, we feel that all of these have made and continue to make a significant impact on society. And we hope you realize that more applications are being discovered almost daily.

Ames Test

The Ames test is a test that is used to determine whether a substance will affect (mutate) the structure of DNA. In this test salmonella bacteria is exposed to the chemical under question (food additives, for example), and changes in the way the bacteria grows are measured. Many substances that cause mutations in this bacteria also cause cancer in animals and humans. Indeed, this test is used today to screen chemicals for their potential ability to cause cancer in humans.

Pregnancy Testing

There are two types of pregnancy tests — one uses a urine sample and the other a blood sample. Both detect the presence of the hormone human chorionic gonadotropin (hCG). This hormone is produced by the placenta shortly after implantation of the embryo into the uterine walls and accumulates

rapidly in the body in the first few days after implantation. Home pregnancy tests, urine tests, are typically around 97 percent accurate if done two weeks after implantation. Blood tests, performed in a clinic, are more costly but can detect pregnancy as early as a week after implantation.

HIV Testing

Tests have been developed to screen for the presence of the human immunodeficiency virus. These tests may be done on urine, blood serum, or saliva and detect HIV antigens, antibodies, or nucleic acids (RNA). The nucleic acid tests (NAT) detect a 142-base sequence located on one of the HIV genes. Most blood banks use a combination of tests to ensure accuracy.

Breast Cancer Testing

Most breast cancer is not hereditary, but in 5–10 percent of cases, there *is* a heredity linkage. The vast majority of these cases is due to mutations in two genes: Breast Cancer-1 gene (BRCA1) and the Breast Cancer-2 gene (BRCA2), which were discovered in 1994 and 1995, respectively. Females who inherit a mutation in either one of these genes have a greatly increased chance of developing breast cancer over their lifetime. Postive tests for these mutations allow the individual to schedule increased screening tests at a more frequent rate than the general population.

Prenatal Genetic Testing

Prenatal genetic testing refers to testing the fetus for potential genetic defects. Tests commonly are performed on blood or tissue samples from the fetus. This may involve amniocentesis — collection of a sample of amniotic fluid that contains cells from the fetus — or collection of blood from the umbilical cord. Tests such as these are used to detect chromosomal abnormalities, such as Down syndrome or birth defects such as spina bifida.

PKU Screening

Phenylketonuria (PKU) is a metabolic disorder in which the individual is missing an enzyme called phenylalanine hydroxylase. Absence of this enzyme allows the buildup of phenylalanine, which can lead to mental retardation.

All states in the United States require PKU testing of all newborns. Infants who test positive are placed on a diet low in phenylalanine, allowing them to develop normally. Check out cans of soft drinks, and you will find a warning on many of them that they contain phenylalanine.

Genetically Modified Foods

Biochemists have developed the ability to transfer genes from one organism into other organisms, including plants and animals. This allows the creation of crops that are more pest and disease resistant and animals that are more disease resistant. Genetic modification can also be used to increase the yield of milk, eggs, or meat. In 1993, the first genetically modified food was given a license for human consumption from the U.S. Food and Drug Administration. It was a new tomato called Flavr Savr, which was resistant to rotting. However, the public has been slow to accept genetically altered foods and afraid of unforeseen effects on the population and environment.

Genetic Engineering

Genetic engineering involves taking a gene from one organism and placing it into another. The recipient may be a bacteria or a plant or an animal. One of the most well-known examples of genetic engineering involves the hormone insulin. Diabetes use to be treated with insulin derived from pigs or cows, and although very similar to human insulin, these animal-derived insulins were not identical and caused problems for some individuals. Biochemists solved this problem by inserting the gene for human insulin into bacteria. The bacteria, through the process of translation, created human insulin. (See Chapter 17 for much more on translation.)

Cloning

In 1996, Dolly the sheep was cloned — the first mammal ever cloned from adult animal cells. The cloned sheep was, of course, genetically identical to the original adult sheep. This clone was created by taking cells from the udder of a 6-year-old ewe and growing them in the lab. They then took unfertilized eggs and stripped the genetic material from them. Finally, they inserted the genetic material grown in the lab into these cells and implanted them into the uterus of another sheep. And Dolly was born. Since Dolly, many other animals have been successfully cloned. However, there is worldwide debate on the idea of cloning a human, which will surely continue for decades.

Gene-Replacement Therapy

In gene-replacement therapy, a modified or healthy gene is inserted into the organism to replace a disease-causing gene. Commonly a virus that has been altered to carry human DNA is used to deliver the healthy gene to the targeted cells of the patient. This process was first used successfully in 1990 on a 4-year-old patient who lacked an immune system due to a rare genetic disease called severe combined immunodeficiency (SCID). Individuals with SCID were prone to life-threatening infections. They lead isolated lives, avoiding people and commonly taking massive doses of antibiotics. Scientists removed white blood cells from the patient, grew them in the lab, and inserted the missing gene into the cells. They then inserted this genetically altered blood back into the patient. The process allowed the child to develop normally and even attend school, but the treatment must be repeated every few months.

Chapter 19

Ten Biochemistry Careers

In This Chapter

- Cruising careers
- Finding out about professions
- Juggling jobs

Because of recent advances in biochemistry and its related area biotechnology, many new professions have been created for the individual majoring in biochemistry. Those who stop at the B.S. degree often find themselves working as technicians in a variety of industries, but for those who go on for their M.S. or Ph.D. many more opportunities become possible.

Graduates at all levels find positions in a wide variety of career areas including forensics, industrial chemistry, molecular biology, pharmacology, technical sales, virology, horticulture, immunology, forestry, and so on. We have mentioned several careers throughout the book, so we are including here careers one might not normally associate with the field of biochemistry.

Research Assistant

A research assistant works in the area of biochemical research and development as part of a team. They may investigate new genetic tests, be involved in genetic engineering or cloning, or help with the development of new drugs or drug protocols. In addition to performing typical technical biochemical procedures, the research assistant analyzes data and prepares technical reports and summaries. Research assistants are often also involved in the search for inventions that can lead to patents. They may eventually head up their own research groups.

Plant Breeder

A plant breeder designs and implements plant breeding projects in conjunction with other research teams. They may be involved in the development of disease-resistant strains of crops or may search for ways to increase crop yields using biochemical and biotechnological techniques. They may also be involved in personnel management, public relations, and/or advising their company about future projects and plant-breeding goals.

Quality Control Analyst

The quality control analyst conducts analyses of raw materials and the finished products coming off the production line. They collect data concerning quality control test procedures and pinpoint sources of error. Along with quality control engineers, the analyst ensures that the quality of the product remains high. This is especially important, as you might imagine, when the product could be a genetically modified virus or a genetically altered food.

Clinical Research Associate

Clinical research associates design and implement clinical research projects such as a new drug protocol or the use of a new virus for gene therapy. They may travel to the various field sites where the clinical trials are being conducted to coordinate and/or supervise the trials. The clinical associate analyzes and evaluates data from the trials to ensure that clinical protocols were followed. A background in nursing or pharmacology is useful.

Technical Writer

Anyone who has ever read a poorly written set of directions or technical manual realizes the importance of a good technical writer. A technical writer in the biochemical world edits and writes operating procedures, laboratory manuals, clinical protocols, and so on. They ensure that these documents are written in a way that meets government regulations. They may develop professional development programs for staff members and write news releases. Part of their job is to take highly technical reports and edit them in such a way that they are understandable to the company's administration and the general public.

Biochemical Development Engineer

The biochemical development engineer takes the biochemical process developed in the laboratory and scales it up through the pilot plant stage to the full production plant. They help determine what instrumentation and equipment are needed and troubleshoot problems in the scale-up procedure. They work to develop more efficient manufacturing processes while maintaining a high degree of quality control. They may also be involved in technological advances from another area and apply them to their manufacturing process.

Market Research Analyst

Market research analysts analyze and research the company's market, the product mix, and the competition. They perform literature searches and make presentations on technical areas and new potential markets for the company. They predict future marketing trends based upon market research and may even be involved in the preparation of research proposals.

Patent Attorney

A patent attorney coordinates and prepares documentation for patent applications. They track the company's research studies and recommend the timing of patent filings. They collect supporting documentation and negotiate patent licenses and other legal agreements. They may become involved with interference and appeal hearings.

Pharmaceutical Sales

An individual with a degree in biochemistry becomes a natural for a career in pharmaceutical sales. These sales representatives spend much of their time on the road, talking to hospital personnel, physicians, pharmacists, and others. They are quite familiar with their company's products and try to be as persuasive as possible in touting their advantages over the competition. They have to be familiar with statistics and issues of concern in the medical community in order to successfully communicate with potential clients.

Biostatistician

Biostatisticians are statisticians who work in health-related fields. They design research studies and collect and analyze data on problems — such as how a disease progresses, how safe a new treatment or medication is, or the impact of certain risk factors associated with medical conditions. They may also design and analyze studies to determine health care costs and health care quality. They are instrumental in the designing stages of studies, providing expertise on experimental design, sample sizes, and other considerations.

Index

• Symbols & Numerics •

• A •

• B •

• C •

• D •

• E •

• F •

• G •

• H •

• I •

• J •

• K •

• L •

• M •

• N •

• O •

• P •

• Q •

• R •

• S •

• T •

• U •

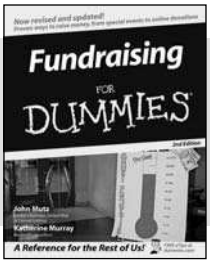
• V •

• W •

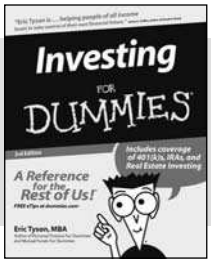
• X •

• Z •

BUSINESS, CAREERS & PERSONAL FINANCE



0-7645-9847-3

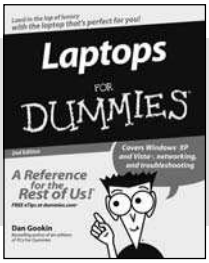


0-7645-2431-3

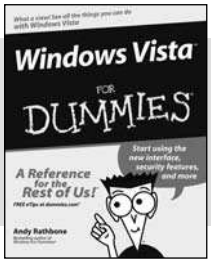
Also available:

- Business Plans Kit For Dummies 0-7645-9794-9
- Economics For Dummies 0-7645-5726-2
- Grant Writing For Dummies 0-7645-8416-2
- Home Buying For Dummies 0-7645-5331-3
- Managing For Dummies 0-7645-1771-6
- Marketing For Dummies 0-7645-5600-2
- Personal Finance For Dummies 0-7645-2590-5*
- Resumes For Dummies 0-7645-5471-9
- Selling For Dummies 0-7645-5363-1
- Six Sigma For Dummies 0-7645-6798-5
- Small Business Kit For Dummies 0-7645-5984-2
- Starting an eBay Business For Dummies 0-7645-6924-4
- Your Dream Career For Dummies 0-7645-9795-7

HOME & BUSINESS COMPUTER BASICS



0-470-05432-8

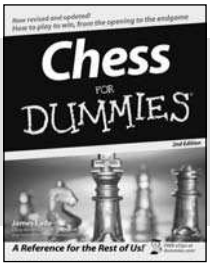


0-471-75421-8

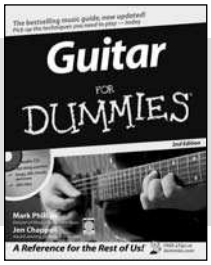
Also available:

- Cleaning Windows Vista For Dummies 0-471-78293-9
- Excel 2007 For Dummies 0-470-03737-7
- Mac OS X Tiger For Dummies 0-7645-7675-5
- MacBook For Dummies 0-470-04859-X
- Macs For Dummies 0-470-04849-2
- Office 2007 For Dummies 0-470-00923-3
- Outlook 2007 For Dummies 0-470-03830-6
- PCs For Dummies 0-7645-8958-X
- Salesforce.com For Dummies 0-470-04893-X
- Upgrading & Fixing Laptops For Dummies 0-7645-8959-8
- Word 2007 For Dummies 0-470-03658-3
- Quicken 2007 For Dummies 0-470-04600-7

FOOD, HOME, GARDEN, HOBBIES, MUSIC & PETS



0-7645-8404-9

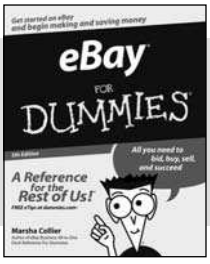


0-7645-9904-6

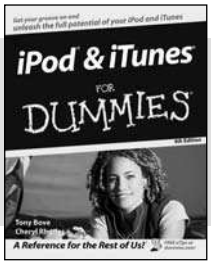
Also available:

- Candy Making For Dummies 0-7645-9734-5
- Card Games For Dummies 0-7645-9910-0
- Crocheting For Dummies 0-7645-4151-X
- Dog Training For Dummies 0-7645-8418-9
- Healthy Carb Cookbook For Dummies 0-7645-8476-6
- Home Maintenance For Dummies 0-7645-5215-5
- Horses For Dummies 0-7645-9797-3
- Jewelry Making & Beading For Dummies 0-7645-2571-9
- Orchids For Dummies 0-7645-6759-4
- Puppies For Dummies 0-7645-5255-4
- Rock Guitar For Dummies 0-7645-5356-9
- Sewing For Dummies 0-7645-6847-7
- Singing For Dummies 0-7645-2475-5

INTERNET & DIGITAL MEDIA



0-470-04529-9



0-470-04894-8

Also available:

- Blogging For Dummies 0-471-77084-1
- Digital Photography For Dummies 0-7645-9802-3
- Digital Photography All-in-One Desk Reference For Dummies 0-470-03743-1
- Digital SLR Cameras and Photography For Dummies 0-7645-9803-1
- eBay Business All-in-One Desk Reference For Dummies 0-7645-8438-3
- HDTV For Dummies 0-470-09673-X
- Home Entertainment PCs For Dummies 0-470-05523-5
- MySpace For Dummies 0-470-09529-6
- Search Engine Optimization For Dummies 0-471-97998-8
- Skype For Dummies 0-470-04891-3
- The Internet For Dummies 0-7645-8996-2
- Wiring Your Digital Home For Dummies 0-471-91830-X

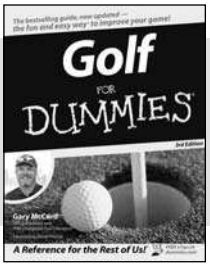
*** Separate Canadian edition also available**

† **Separate U.K. edition also available**

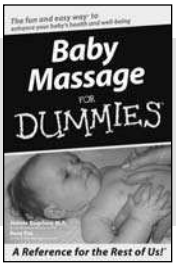
Available wherever books are sold. For more information or to order direct: U.S. customers visit www.dummies.com or call 1-877-762-2974. U.K. customers visit www.wileyeurope.com or call 0800 243407. Canadian customers visit www.wiley.ca or call 1-800-567-4797.

WILEY

SPORTS, FITNESS, PARENTING, RELIGION & SPIRITUALITY



0-471-76871-5

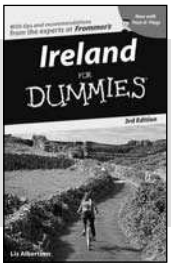


0-7645-7841-3

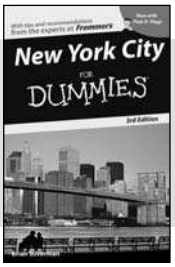
Also available:

- Catholicism For Dummies 0-7645-5391-7
- Exercise Balls For Dummies 0-7645-5623-1
- Fitness For Dummies 0-7645-7851-0
- Football For Dummies 0-7645-3936-1
- Judaism For Dummies 0-7645-5299-6
- Potty Training For Dummies 0-7645-5417-4
- Buddhism For Dummies 0-7645-5359-3
- Pregnancy For Dummies 0-7645-4483-7 †
- Ten Minute Tone-Ups For Dummies 0-7645-7207-5
- NASCAR For Dummies 0-7645-7681-X
- Religion For Dummies 0-7645-5264-3
- Soccer For Dummies 0-7645-5229-5
- Women in the Bible For Dummies 0-7645-8475-8

TRAVEL



0-7645-7749-2

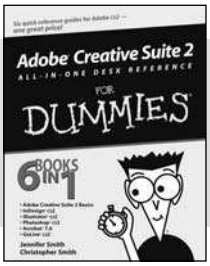


0-7645-6945-7

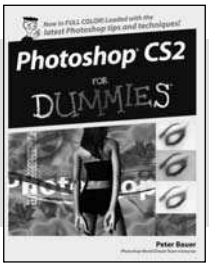
Also available:

- Alaska For Dummies 0-7645-7746-8
- Cruise Vacations For Dummies 0-7645-6941-4
- England For Dummies 0-7645-4276-1
- Europe For Dummies 0-7645-7529-5
- Germany For Dummies 0-7645-7823-5
- Hawaii For Dummies 0-7645-7402-7
- Italy For Dummies 0-7645-7386-1
- Las Vegas For Dummies 0-7645-7382-9
- London For Dummies 0-7645-4277-X
- Paris For Dummies 0-7645-7630-5
- RV Vacations For Dummies 0-7645-4442-X
- Walt Disney World & Orlando For Dummies 0-7645-9660-8

GRAPHICS, DESIGN & WEB DEVELOPMENT



0-7645-8815-X

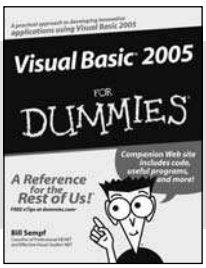


0-7645-9571-7

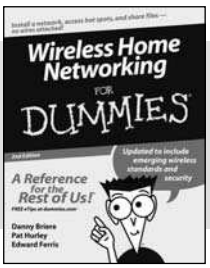
Also available:

- 3D Game Animation For Dummies 0-7645-8789-7
- AutoCAD 2006 For Dummies 0-7645-8925-3
- Building a Web Site For Dummies 0-7645-7144-3
- Creating Web Pages For Dummies 0-470-08030-2
- Creating Web Pages All-in-One Desk Reference For Dummies 0-7645-4345-8
- Dreamweaver 8 For Dummies 0-7645-9649-7
- InDesign CS2 For Dummies 0-7645-9572-5
- Macromedia Flash 8 For Dummies 0-7645-9691-8
- Photoshop CS2 and Digital Photography For Dummies 0-7645-9580-6
- Photoshop Elements 4 For Dummies 0-471-77483-9
- Syndicating Web Sites with RSS Feeds For Dummies 0-7645-8848-6
- Yahoo! SiteBuilder For Dummies 0-7645-9800-7

NETWORKING, SECURITY, PROGRAMMING & DATABASES



0-7645-7728-X



0-471-74940-0

Also available:

- Access 2007 For Dummies 0-470-04612-0
- ASP.NET 2 For Dummies 0-7645-7907-X
- C# 2005 For Dummies 0-7645-9704-3
- Hacking For Dummies 0-470-05235-X
- Hacking Wireless Networks For Dummies 0-7645-9730-2
- Java For Dummies 0-470-08716-1
- Microsoft SQL Server 2005 For Dummies 0-7645-7755-7
- Networking All-in-One Desk Reference For Dummies 0-7645-9939-9
- Preventing Identity Theft For Dummies 0-7645-7336-5
- Telecom For Dummies 0-471-77085-X
- Visual Studio 2005 All-in-One Desk Reference For Dummies 0-7645-9775-2
- XML For Dummies 0-7645-8845-1