CHAPTER

16

Soap bubbles, so common yet so beautiful, are made from animal fat, a lipid.

Biomolecules: Lipids and Nucleic Acids

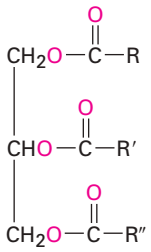
In the previous two chapters, we've discussed the organic chemistry of carbohydrates and proteins, two of the four major classes of biomolecules. Let's now look at the two remaining classes, beginning with *lipids* and continuing on to *nucleic acids*.

Lipids are naturally occurring molecules that have limited solubility in water and can be isolated from organisms by extraction with a nonpolar organic solvent. Fats, oils, waxes, many vitamins and hormones, and most nonprotein cell-membrane components are examples. Note that, unlike carbohydrates and proteins, lipids are defined by a physical property (solubility) rather than by structure.

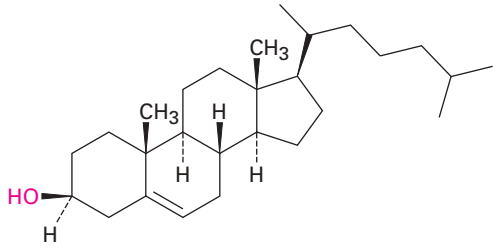
Lipids are classified into two general types: those like fats and waxes, which contain ester linkages and can be hydrolyzed, and those like cholesterol and other steroids, which don't have ester linkages and can't be hydrolyzed.

OWL

Online homework for this chapter can be assigned in OWL, an online homework assessment tool.



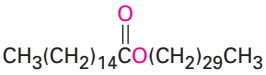
Animal fat—a triester
(R, R′, R″ = C_{11}–C_{19} chains)



Cholesterol

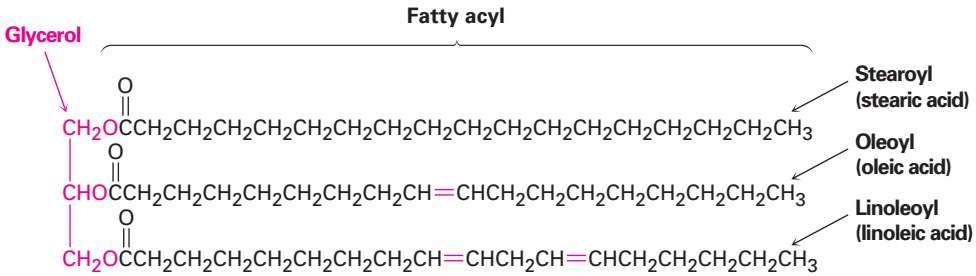
16.1 Waxes, Fats, and Oils

Waxes are mixtures of esters of long-chain carboxylic acids with long-chain alcohols. The carboxylic acid usually has an even number of carbons from 16 through 36, while the alcohol has an even number of carbons from 24 through 36. One of the major components of beeswax, for instance, is triacontyl hexadecanoate, the ester of the C_{30} alcohol triacontan-1-ol and the C_{16} acid hexadecanoic acid. The waxy protective coatings on most fruits, berries, leaves, and animal furs have similar structures.



Triacontyl hexadecanoate (from beeswax)

Animal fats and vegetable oils are the most widely occurring lipids. Although they appear different—animal fats like butter and lard are solids, whereas vegetable oils like corn oil and peanut oil are liquids—their structures are closely related. Fats and oils are *triglycerides,* or **triacylglycerols**—triesters of glycerol with three long-chain carboxylic acids called **fatty acids**. Animals use fats for long-term energy storage because they are much less highly oxidized than carbohydrates and provide about six times as much energy as an equal weight of stored, hydrated glycogen.



A triacylglycerol

Hydrolysis of a fat or oil with aqueous NaOH yields glycerol and three fatty acids. The fatty acids are generally unbranched and contain an even

number of carbon atoms between 12 and 20. If double bonds are present, they have largely, although not entirely, *Z,* or cis, geometry. The three fatty acids of a specific triacylglycerol molecule need not be the same, and the fat or oil from a given source is likely to be a complex mixture of many different triacylglycerols. Table 16.1 lists some of the commonly occurring fatty acids, and Table 16.2 lists the approximate composition of some fats and oils from different sources.

Table 16.1 Structures of Some Common Fatty Acids

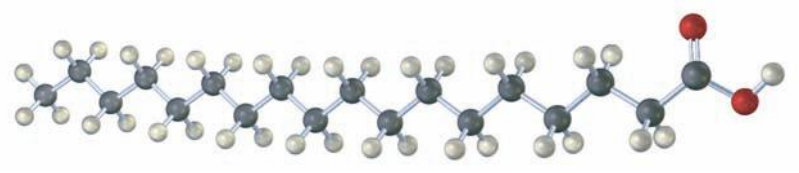
Name	No. of carbons	Melting point (°C)	Structure
Saturated			
Lauric	12	43.2	$CH_3(CH_2)_{10}CO_2H$
Myristic	14	53.9	$CH_3(CH_2)_{12}CO_2H$
Palmitic	16	63.1	$CH_3(CH_2)_{14}CO_2H$
Stearic	18	68.8	$CH_3(CH_2)_{16}CO_2H$
Arachidic	20	76.5	$CH_3(CH_2)_{18}CO_2H$
Unsaturated			
Palmitoleic	16	−0.1	(Z)-$CH_3(CH_2)_5CH=CH(CH_2)_7CO_2H$
Oleic	18	13.4	(Z)-$CH_3(CH_2)_7CH=CH(CH_2)_7CO_2H$
Linoleic	18	−12	(Z,Z)-$CH_3(CH_2)_4(CH=CHCH_2)_2(CH_2)_6CO_2H$
Linolenic	18	−11	(all Z)-$CH_3CH_2(CH=CHCH_2)_3(CH_2)_6CO_2H$
Arachidonic	20	−49.5	(all Z)-$CH_3(CH_2)_4(CH=CHCH_2)_4CH_2CH_2CO_2H$

Table 16.2 Composition of Some Fats and Oils

	Saturated fatty acids (%)				Unsaturated fatty acids (%)	
Source	C_{12} lauric	C_{14} myristic	C_{16} palmitic	C_{18} stearic	C_{18} oleic	C_{18} linoleic
Animal fat						
Lard	—	1	25	15	50	6
Butter	2	10	25	10	25	5
Human fat	1	3	25	8	46	10
Whale blubber	—	8	12	3	35	10
Vegetable oil						
Coconut	50	18	8	2	6	1
Corn	—	1	10	4	35	45
Olive	—	1	5	5	80	7
Peanut	—	—	7	5	60	20

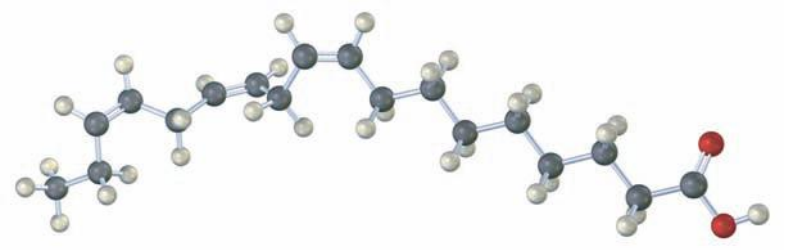
More than 100 different fatty acids are known, and about 40 occur widely. Palmitic acid (C_{16}) and stearic acid (C_{18}) are the most abundant saturated fatty acids; oleic and linoleic acids (both C_{18}) are the most abundant unsaturated ones. Oleic acid is *monounsaturated* because it has only one double

bond, whereas linoleic, linolenic, and arachidonic acids are **polyunsaturated fatty acids** because they have more than one double bond. Linoleic and linolenic acids occur in cream and are essential in the human diet; infants grow poorly and develop skin lesions if fed a diet of nonfat milk for prolonged periods.



$$CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2CH_2C(=O)OH$$

Stearic acid



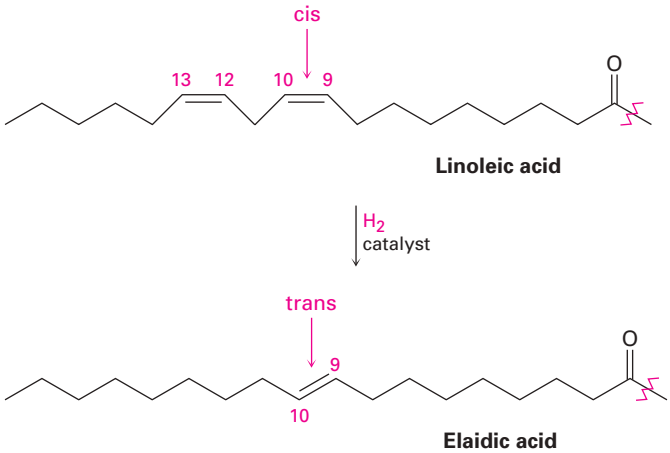
$$CH_3CH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CHCH_2CH_2CH_2CH_2CH_2CH_2CH_2C(=O)OH$$

Linolenic acid, a polyunsaturated fatty acid

The data in Table 16.1 show that unsaturated fatty acids generally have lower melting points than their saturated counterparts, a trend that is also true for triacylglycerols. Since vegetable oils generally have a higher proportion of unsaturated to saturated fatty acids than animal fats (Table 16.2), they have lower melting points. The difference is a consequence of structure. Saturated fats have a uniform shape that allows them to pack together efficiently in a crystal lattice. In unsaturated vegetable oils, however, the C=C bonds introduce bends and kinks into the hydrocarbon chains, making crystal formation less favorable and lowering the melting point.

The C=C bonds in vegetable oils can be reduced by catalytic hydrogenation, typically carried out at high temperature using a nickel catalyst, to produce saturated solid or semisolid fats. Margarine and shortening are produced by hydrogenating soybean, peanut, or cottonseed oil until the proper consistency is obtained. Unfortunately, the hydrogenation reaction is accompanied by some cis–trans isomerization of the double bonds that remain, producing fats with about 10% to 15% trans unsaturated fatty acids. Dietary intake of trans fatty acids increases cholesterol levels in the blood,

thereby increasing the risk for heart problems. The conversion of linoleic acid into elaidic acid is an example.

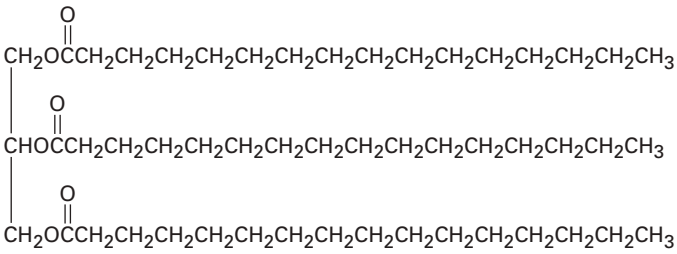


Worked Example 16.1 Drawing the Structure of a Fat

Draw the structure of glyceryl tripalmitate, a typical fat molecule.

Strategy As the name implies, glyceryl tripalmitate is the triester of glycerol with three molecules of palmitic acid, $CH_3(CH_2)_{14}CO_2H$.

Solution



Glyceryl tripalmitate

Problem 16.1 Carnauba wax, used in floor and furniture polishes, contains an ester of a C_{32} straight-chain alcohol with a C_{20} straight-chain carboxylic acid. Draw its structure.

Problem 16.2 Draw structures of the following compounds. Which would you expect to have a higher melting point?

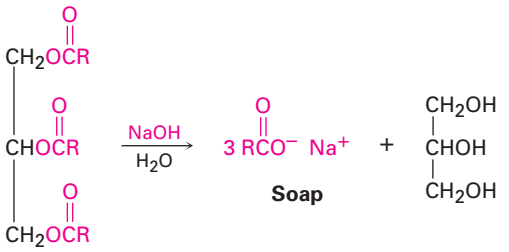
(a) Glyceryl trioleate **(b)** Glyceryl monooleate distearate

Problem 16.3 Fats and oils can be either optically active or optically inactive, depending on their structures. Draw the structure of an optically active fat that gives 2 equivalents of palmitic acid and 1 equivalent of stearic acid on hydrolysis. Draw the structure of an optically inactive fat that gives the same products on hydrolysis.

16.2 Soaps

Soap has been known since at least 600 BC, when the Phoenicians prepared a curdy material by boiling goat fat with extracts of wood ash. The cleansing properties of soap weren't generally recognized, however, and the use of soap

didn't become widespread until the 18th century. Chemically, soap is a mixture of the sodium or potassium salts of long-chain fatty acids produced by hydrolysis (*saponification*) of animal fat with alkali.



A fat
(R = C_{11}–C_{19} aliphatic chains)

Glycerol

Crude soap curds contain glycerol and excess alkali as well as soap but can be purified by boiling with water and adding NaCl or KCl to precipitate the pure carboxylate salts. The smooth soap that precipitates is dried, perfumed, and pressed into bars for household use. Dyes are added to make colored soaps, antiseptics are added for medicated soaps, pumice is added for scouring soaps, and air is blown in for soaps that float.

Soaps act as cleansers because the two ends of a soap molecule are so different. The carboxylate end of the long-chain molecule is ionic and therefore *hydrophilic,* or attracted to water. The long hydrocarbon portion of the molecule, however, is nonpolar and *hydrophobic,* or water avoiding, and therefore more soluble in oils. The net effect of these two opposing tendencies is that soaps are attracted to both oils and water.

When soaps are dispersed in water, the long hydrocarbon tails cluster together on the inside of tangled, hydrophobic balls, while the ionic heads on the surface of the clusters stick out into the water layer. These spherical clusters, called **micelles**, are shown schematically in Figure 16.1. Grease and oil droplets are solubilized in water when they are coated by the nonpolar, hydrophobic tails of soap molecules in the center of micelles. Once solubilized, the grease and dirt can be rinsed away.

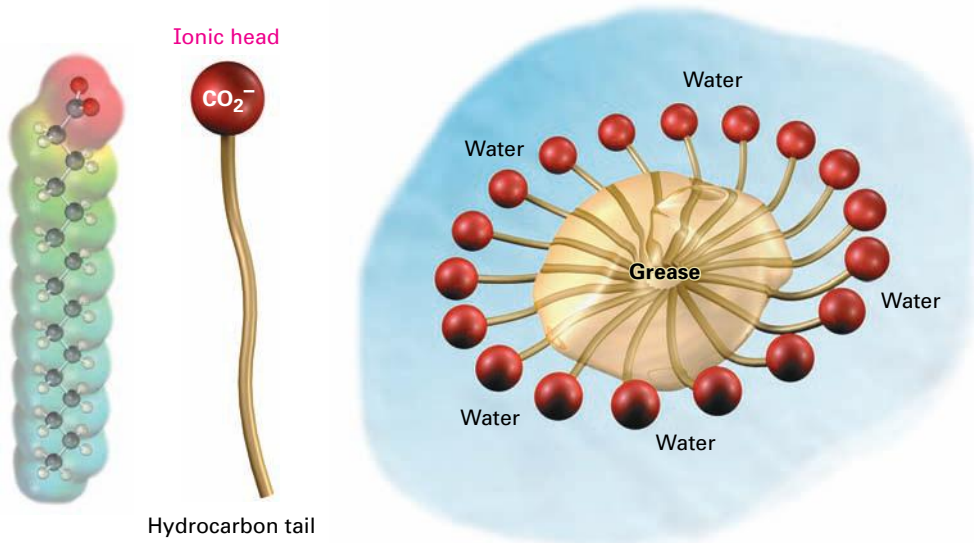
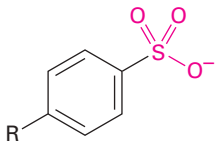


Figure 16.1 A soap micelle solubilizing a grease particle in water. An electrostatic potential map of a fatty-acid carboxylate shows how the negative charge is located in the head group.

As useful as they are, soaps also have some drawbacks. In hard water, which contains metal ions such as Mg^{2+}, Ca^{2+}, and Fe^{3+}, soluble sodium carboxylates are converted into insoluble metal salts, leaving the familiar ring of scum around bathtubs and the gray tinge on white clothes. Chemists have circumvented these problems by synthesizing a class of synthetic detergents based on salts of long-chain alkylbenzenesulfonic acids. The principle of synthetic detergents is the same as that of soaps: the alkylbenzene end of the molecule is attracted to grease, while the anionic sulfonate end is attracted to water. Unlike soaps, though, sulfonate detergents don't form insoluble metal salts in hard water and don't leave an unpleasant scum.



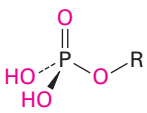
A synthetic detergent
(R = a mixture of C_{12} chains)

Problem 16.4 Draw the structure of magnesium oleate, one of the components of bathtub scum.

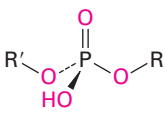
Problem 16.5 Write the saponification reaction of glyceryl dioleate monopalmitate with aqueous NaOH.

16.3 Phospholipids

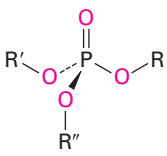
Just as waxes, fats, and oils are esters of carboxylic acids, **phospholipids** are esters of phosphoric acid, H_3PO_4.



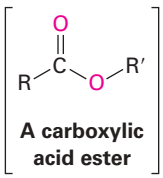
A phosphoric acid monoester



A phosphoric acid diester



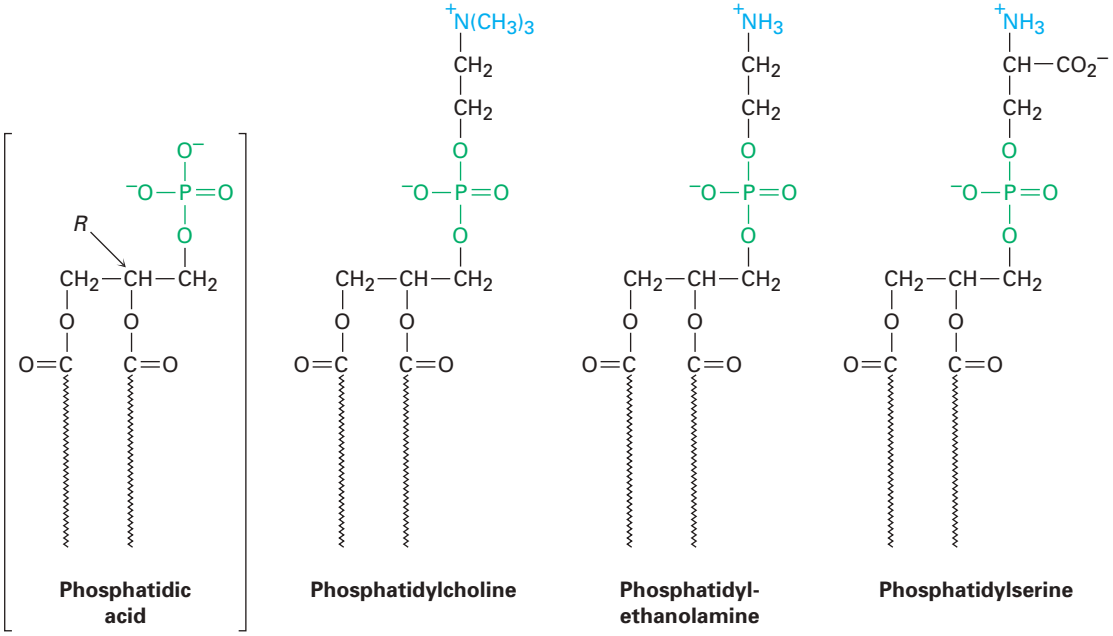
A phosphoric acid triester



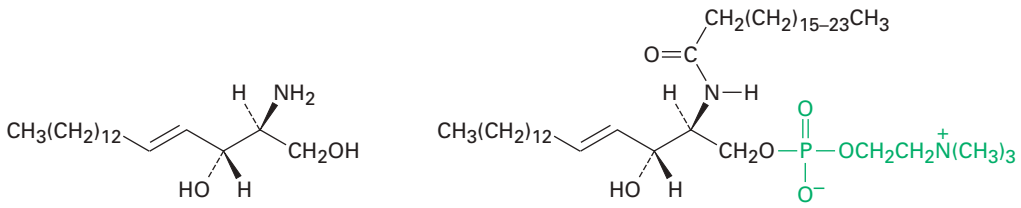
A carboxylic acid ester

Phospholipids are of two general kinds: *glycerophospholipids* and *sphingomyelins*. Glycerophospholipids are based on phosphatidic acid, which contains a glycerol backbone linked by ester bonds to two fatty acids and one phosphoric acid. Although the fatty-acid residues can be any of the C_{12}–C_{20} units

typically present in fats, the acyl group at C1 is usually saturated and the one at C2 is usually unsaturated. The phosphate group at C3 is also bonded to an amino alcohol such as choline $[HOCH_2CH_2N(CH_3)_3]^+$, ethanolamine ($HOCH_2CH_2NH_2$), or serine $[HOCH_2CH(NH_2)CO_2H]$. The compounds are chiral and have an L, or *R,* configuration at C2.



Sphingomyelins are the second major group of phospholipids. These compounds have sphingosine or a related dihydroxyamine as their backbone and are particularly abundant in brain and nerve tissue, where they are a major constituent of the coating around nerve fibers.



Phospholipids are found widely in both plant and animal tissues and make up approximately 50% to 60% of cell membranes. Because they are like soaps in having a long, nonpolar hydrocarbon tail bound to a polar ionic head, phospholipids in the cell membrane organize into a **lipid bilayer** about 5.0 nm (50 Å) thick. As shown in Figure 16.2, the nonpolar tails aggregate in the center of the bilayer in much the same way that soap tails aggregate in the center

of a micelle. This bilayer serves as an effective barrier to the passage of water, ions, and other components into and out of cells.

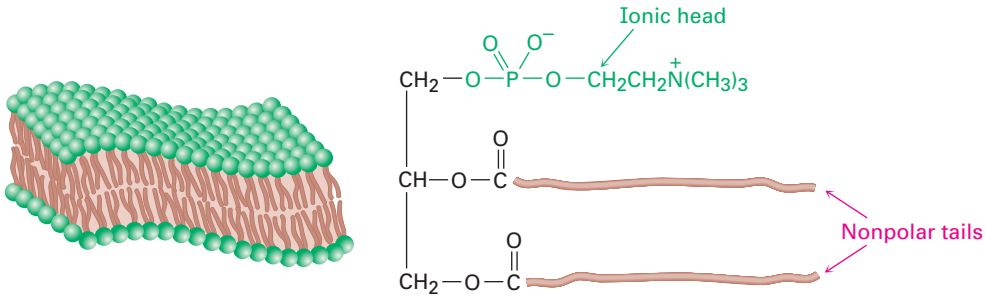
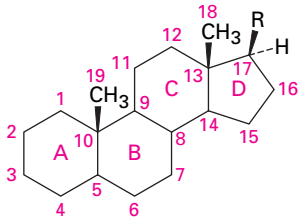


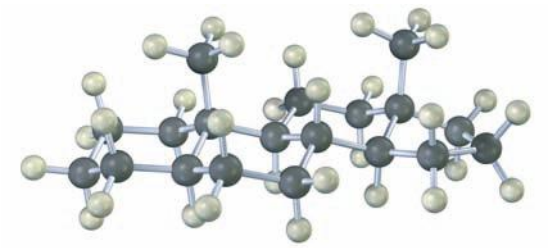
Figure 16.2 Aggregation of glycerophospholipids into the lipid bilayer that composes cell membranes.

16.4 Steroids

In addition to fats and phospholipids, the lipid extracts of plants and animals also contain **steroids**, molecules whose structures are based on a tetracyclic ring system. The four rings are designated A, B, C, and D, beginning at the lower left, and the carbon atoms are numbered beginning in the A ring. The three 6-membered rings (A, B, and C) adopt chair conformations but are constrained by their rigid geometry from undergoing the usual cyclohexane ring-flips (Section 2.11).



A steroid
(R = various side chains)

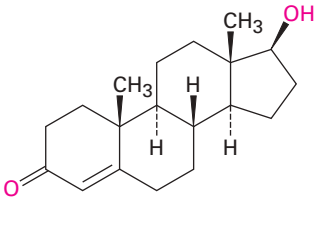


In humans, most steroids function as *hormones,* chemical messengers that are secreted by endocrine glands and carried through the bloodstream to target tissues. There are two main classes of steroid hormones: the *sex hormones,* which control maturation, tissue growth, and reproduction, and the *adrenocortical hormones,* which regulate a variety of metabolic processes.

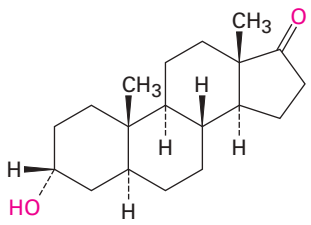
Sex Hormones

Testosterone and androsterone are the two most important male sex hormones, or *androgens.* Androgens are responsible for the development of male

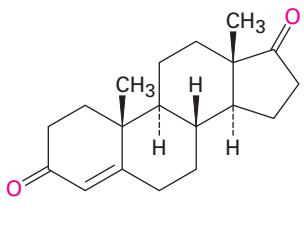
secondary sex characteristics during puberty and for promoting tissue and muscle growth. Both are synthesized in the testes from cholesterol. Androstenedione is another minor hormone that has received particular attention because of its use by prominent athletes.



Testosterone



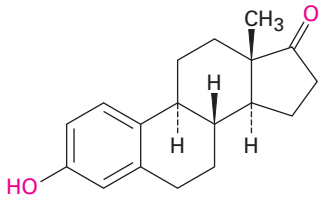
Androsterone



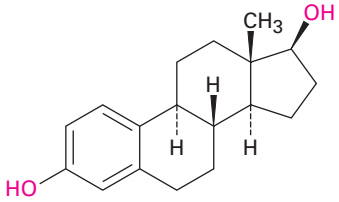
Androstenedione

(Androgens)

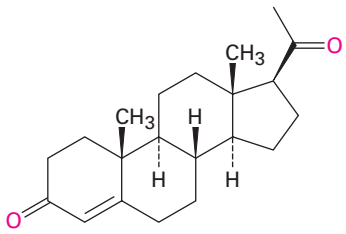
Estrone and estradiol are the two most important female sex hormones, or *estrogens*. Synthesized in the ovaries from testosterone, estrogenic hormones are responsible for the development of female secondary sex characteristics and for regulation of the menstrual cycle. Note that both have a benzene-like aromatic A ring. In addition, another kind of sex hormone, called a *progestin,* is essential for preparing the uterus for implantation of a fertilized ovum during pregnancy. *Progesterone* is the most important progestin.



Estrone



Estradiol



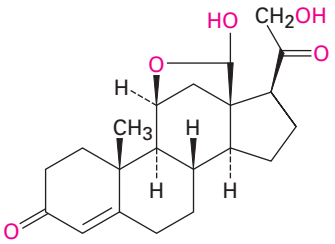
Progesterone (a progestin)

(Estrogens)

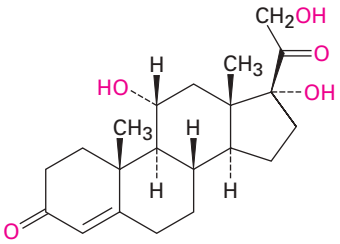
Adrenocortical Hormones

Adrenocortical steroids are secreted by the adrenal glands, small organs located near the upper end of each kidney. There are two types of adrenocortical steroids, called *mineralocorticoids* and *glucocorticoids*. Mineralocorticoids, such as aldosterone, control tissue swelling by regulating cellular salt balance between Na^+ and K^+. Glucocorticoids, such as hydrocortisone, are involved in the regulation of glucose metabolism and in the control of

inflammation. Glucocorticoid ointments are widely used to bring down the swelling from exposure to poison oak or poison ivy.



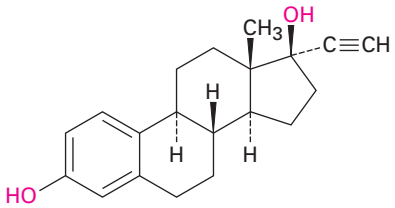
Aldosterone
(a mineralocorticoid)



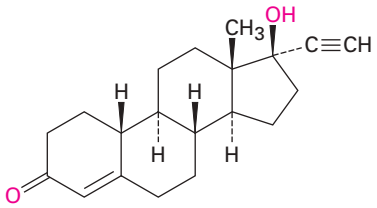
Hydrocortisone
(a glucocorticoid)

Synthetic Steroids

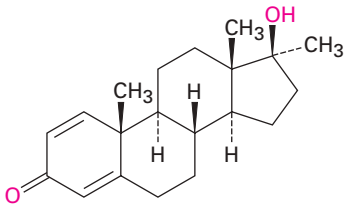
In addition to the many hundreds of steroids isolated from plants and animals, thousands more have been synthesized in pharmaceutical laboratories in a search for new drugs. Among the best-known synthetic steroids are oral contraceptives and anabolic agents. Most birth-control pills are a mixture of two compounds, a synthetic estrogen, such as ethynylestradiol, and a synthetic progestin, such as norethindrone. Anabolic steroids, such as methandrostenolone (Dianabol), are synthetic androgens that mimic the tissue-building effects of natural testosterone.



Ethynylestradiol
(a synthetic estrogen)



Norethindrone
(a synthetic progestin)



Methandrostenolone
(Dianabol)

Problem 16.6 Look at the structure of progesterone, and identify the functional groups in the molecule.

Problem 16.7 Look at the structures of estradiol and ethynylestradiol, and point out the differences. What common structural feature do they share that makes both estrogens?

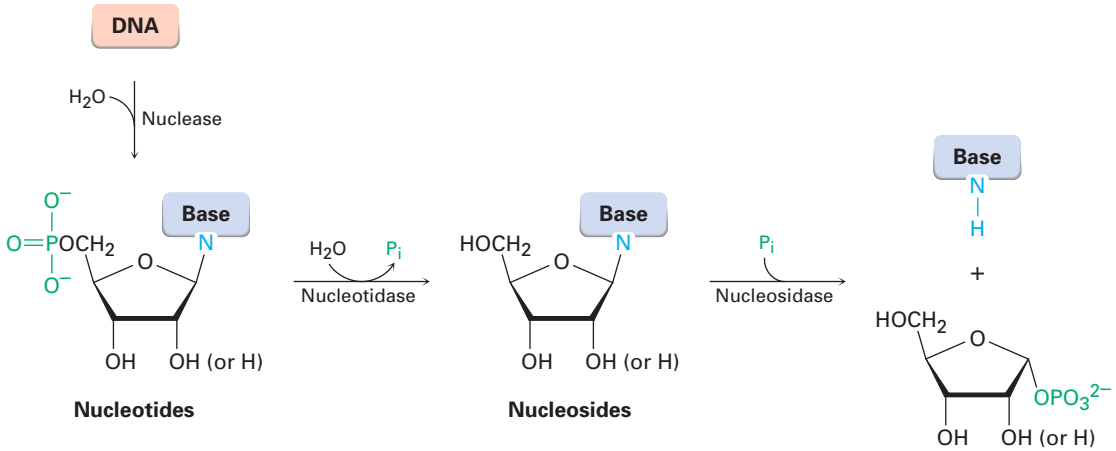
16.5 Nucleic Acids and Nucleotides

Nucleic acids are the last of the four major classes of biomolecules we'll consider. So much has been written and spoken about DNA in the media that you probably know the basics of DNA replication and transcription. The field is moving very rapidly, however, so there's probably a lot you may not be familiar with.

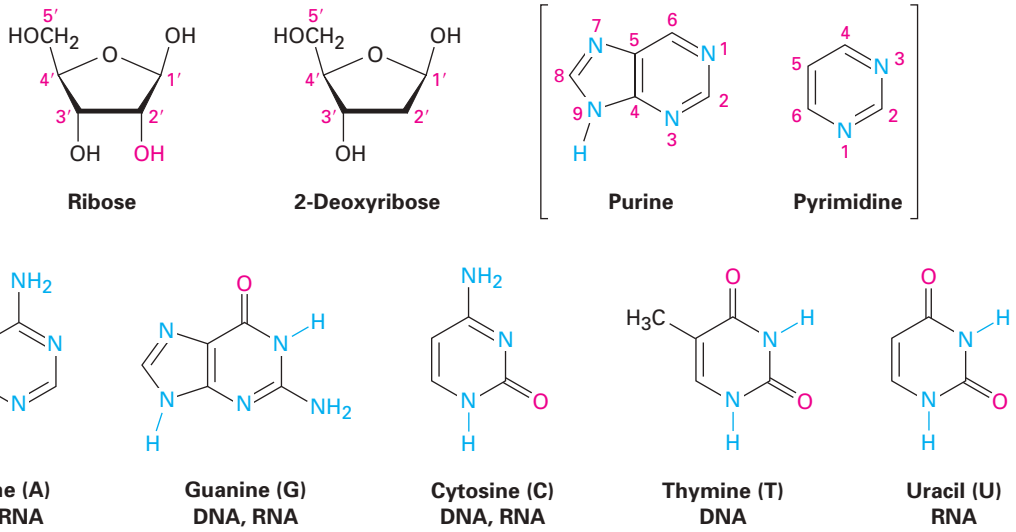
The **nucleic acids**, **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**, are the carriers and processors of a cell's genetic information. Coded in a cell's DNA is all the information that determines the nature of the cell, controls the cell's growth and division, and directs biosynthesis of the enzymes and other

proteins required for cellular functions. In addition to nucleic acids themselves, nucleic acid derivatives such as ATP are involved in many biochemical pathways, and several important coenzymes, including NAD^+, FAD, and coenzyme A, have nucleic acid components. See Table 15.3 on pages 526 and 527 for the structures.

Just as proteins are biopolymers made of amino acids, nucleic acids are biopolymers made of **nucleotides** joined together to form a long chain. Each nucleotide is composed of a **nucleoside** bonded to a phosphate group, and each nucleoside is composed of an aldopentose sugar linked through its anomeric carbon to the nitrogen atom of a heterocyclic amine base (Section 12.6).



The sugar component in RNA is ribose, and the sugar in DNA is 2′-deoxyribose. (In naming and numbering nucleotides, numbers with a prime superscript refer to positions on the sugar, and numbers without a prime superscript refer to positions on the heterocyclic base. Thus, the prefix 2′-deoxy indicates that oxygen is missing from C2′ of ribose.) DNA contains four different amine bases, two substituted purines (adenine and guanine) and two substituted pyrimidines (cytosine and thymine). Adenine, guanine, and cytosine also occur in RNA, but thymine is replaced in RNA by a closely related pyrimidine base called uracil.



The structures of the four deoxyribonucleotides and the four ribonucleotides are shown in Figure 16.3. Although similar chemically, DNA and RNA

differ dramatically in size. Molecules of DNA are enormous, containing as many as 245 million nucleotides and having molecular weights as high as 75 billion. Molecules of RNA, by contrast, are much smaller, containing as few as 21 nucleotides and having molecular weights as low as 7000.

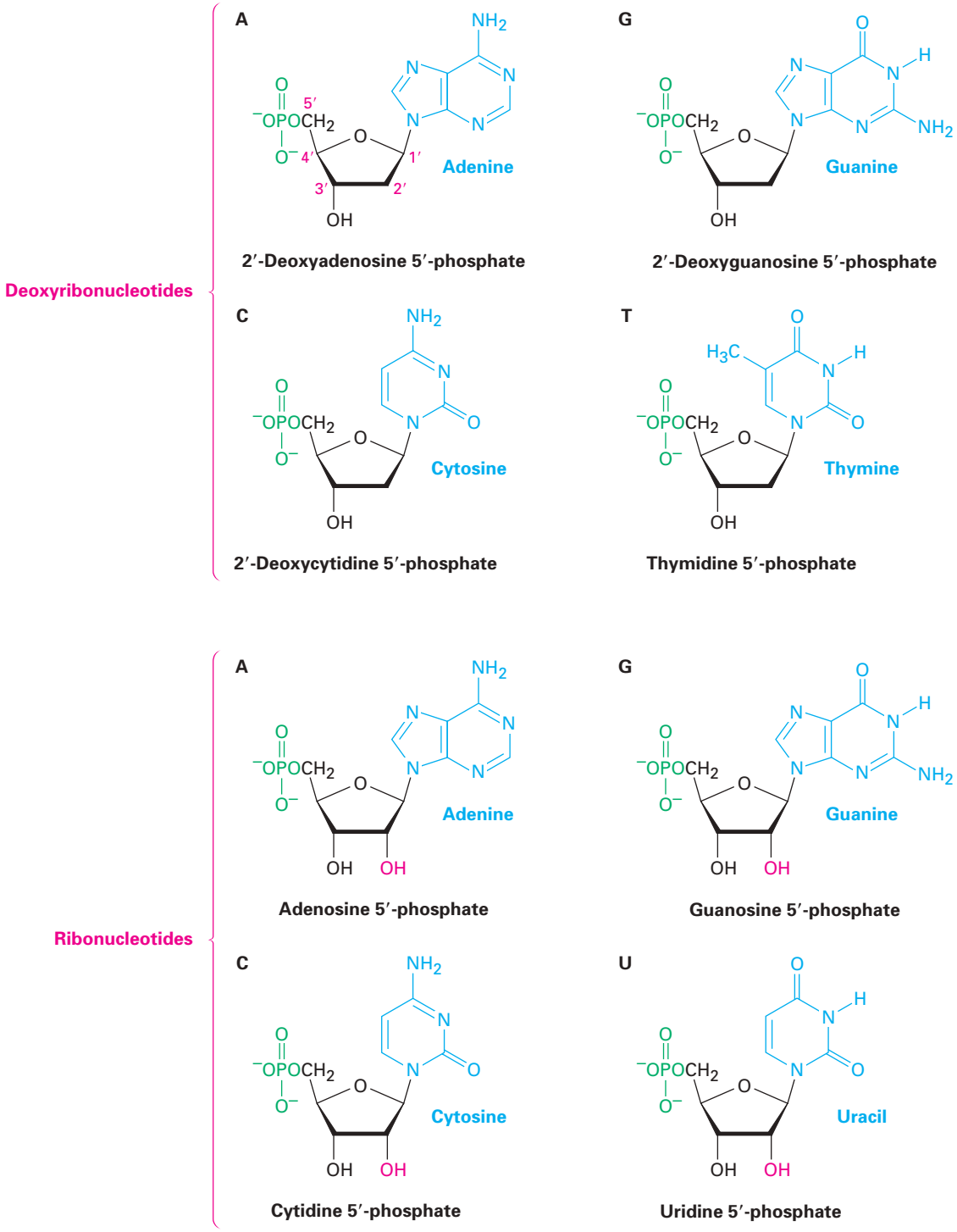
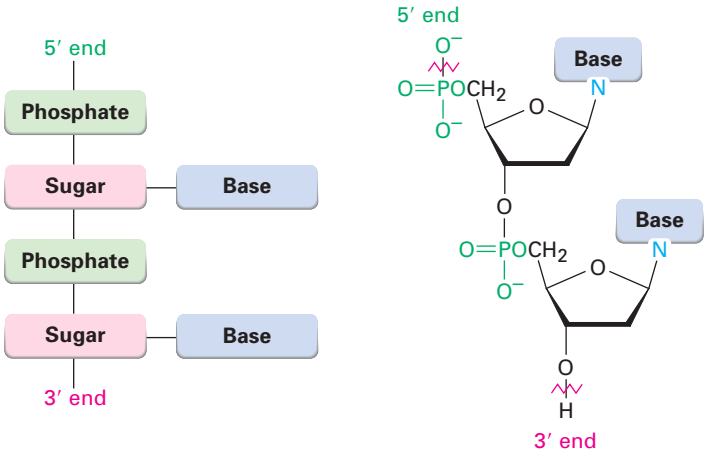


Figure 16.3 Structures of the four deoxyribonucleotides and the four ribonucleotides.

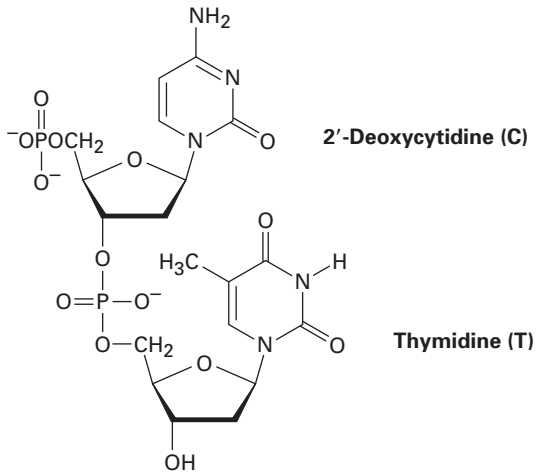
Nucleotides are linked together in DNA and RNA by *phosphodiester* bonds [RO— (PO_2^-)—OR′] between phosphate, the 5′ hydroxyl group on one nucleoside, and the 3′-hydroxyl group on another nucleoside. One end of the nucleic acid polymer has a free hydroxyl at C3′ (the *3′ end*), and the other end has a phosphate at C5′ (the *5′ end*). The sequence of nucleotides in a chain is described by starting at the 5′ end and identifying the bases in order of occurrence, using the abbreviations G, C, A, T (or U for RNA). Thus, a typical DNA sequence might be written as TAGGCT.



Worked Example 16.2 Drawing the Structure of a Dinucleotide

Draw the full structure of the DNA dinucleotide CT.

Solution



Problem 16.8 Draw the full structure of the DNA dinucleotide AG.

Problem 16.9 Draw the full structure of the RNA dinucleotide UA.

16.6 Base Pairing in DNA: The Watson–Crick Model

Samples of DNA isolated from different tissues of the same species have the same proportions of heterocyclic bases, but samples from different species often have greatly different proportions of bases. Human DNA, for example, contains about 30% each of A and T and about 20% each of G and C. The bacterium *Clostridium perfringens,* however, contains about 37% each of A and T and only 13% each of G and C. Note that in both examples, the bases occur in pairs: A and T are present in equal amounts, as are G and C. Why?

In 1953, James Watson and Francis Crick made their historic proposal that DNA consists of two polynucleotide strands, running in opposite directions and coiled around each other in a **double helix** like the handrails on a spiral staircase. The two strands are complementary rather than identical and are held together by hydrogen bonds between specific pairs of bases, A with T and C with G. That is, whenever an A base occurs in one strand, a T base occurs opposite it in the other strand; when a C base occurs in one, a G occurs in the other (Figure 16.4). This complementary base pairing thus explains why A and T are always found in equal amounts, as are G and C.

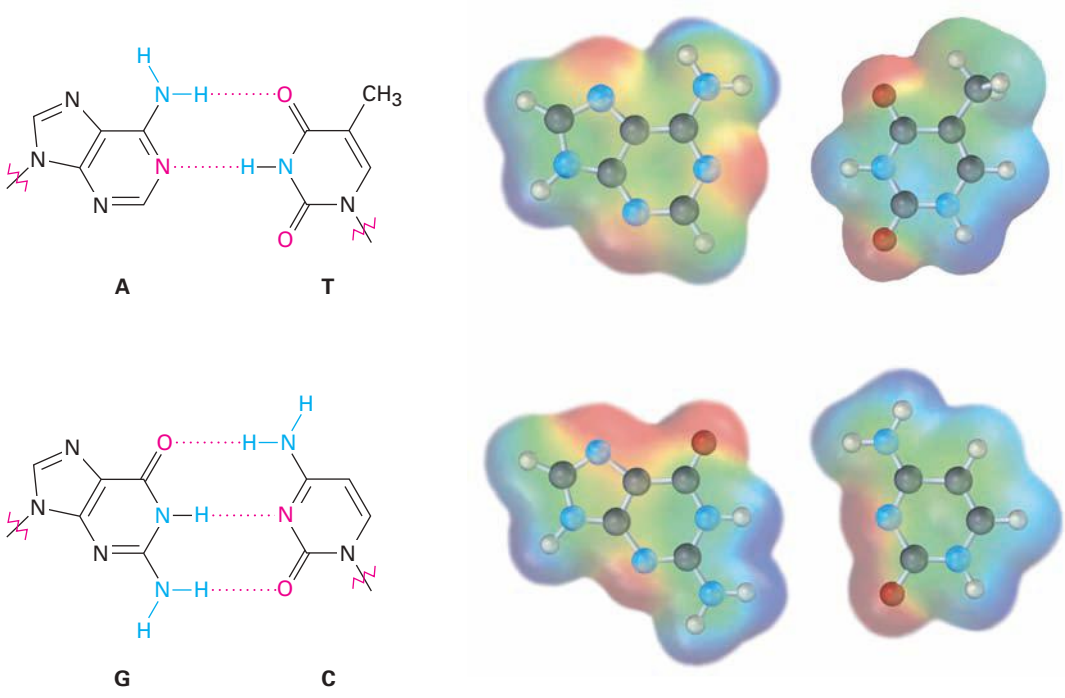


Figure 16.4 Hydrogen-bonding between base pairs in the DNA double helix. Electrostatic potential maps show that the faces of the bases are relatively neutral (green), while the edges have electron-poor (blue) and electron-rich (red) regions. Pairing G with C and A with T brings together the oppositely charged regions.

A full turn of the DNA double helix is shown in Figure 16.5. The helix is 20 Å wide, there are 10 base pairs per turn, and each turn is 34 Å in length. Notice that the two strands of the double helix coil in such a way that two kinds of "grooves" result, a *major groove* 12 Å wide and a *minor groove* 6 Å wide. The major groove is slightly deeper than the minor groove, and both are lined by hydrogen bond donors and acceptors. As a result, a variety of flat, polycyclic aromatic molecules are able to slip sideways, or *intercalate,* between

the stacked bases. Many cancer-causing and cancer-preventing agents function by interacting with DNA in this way.

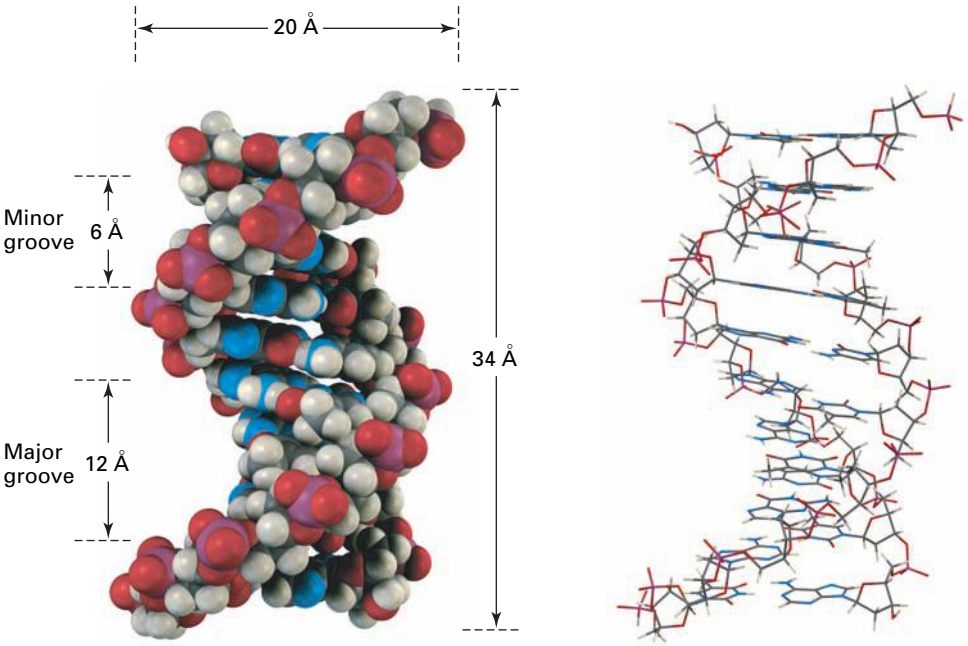
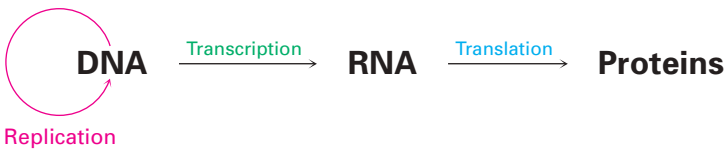


Figure 16.5 A turn of the DNA double helix in both space-filling and wire-frame formats. The sugar–phosphate backbone runs along the outside of the helix, and the amine bases hydrogen bond to one another on the inside. Both major and minor grooves are visible.

An organism's genetic information is stored as a sequence of deoxyribonucleotides strung together in the DNA chain. For the information to be preserved and passed on to future generations, a mechanism must exist for copying DNA. For the information to be used, a mechanism must exist for decoding the DNA message and implementing the instructions it contains. Three fundamental processes take place:

- **Replication**—the process by which identical copies of DNA are made so that information can be preserved and handed down to offspring.
- **Transcription**—the process by which the genetic messages are read and carried out of the cell nucleus to ribosomes, where protein synthesis occurs.
- **Translation**—the process by which the genetic messages are decoded and used to synthesize proteins.



Worked Example 16.3 Complementary DNA Sequences

What sequence of bases on one strand of DNA is complementary to the sequence TATGCAT on another strand?

Strategy Remember that A and G form complementary pairs with T and C, respectively, and then go through the sequence replacing A by T, G by C, T by A, and C by G.

Solution

Original: (5′) TATGCAT (3′)
Complement: (3′) ATACGTA (5′) or (5′) ATGCATA (3′)

Problem 16.10 What sequence of bases on one strand of DNA is complementary to the following sequence on another strand?

(5′) GGCTAATCCGT (3′)

16.7 Replication of DNA

DNA **replication** is an enzyme-catalyzed process that begins with a partial unwinding of the double helix at various points along the chain, brought about by enzymes called *helicases*. Hydrogen bonds are broken, the two strands separate to form a "bubble," and bases are exposed. New nucleotides then line up on each strand in a complementary manner, A to T and G to C, and two new strands begin to grow from the ends of the bubble, called the *replication forks*. Each new strand is complementary to its old template strand, so two identical DNA double helices are produced (Figure 16.6). Because each of the new DNA molecules contains one old strand and one new strand, the process is described as *semiconservative replication*.

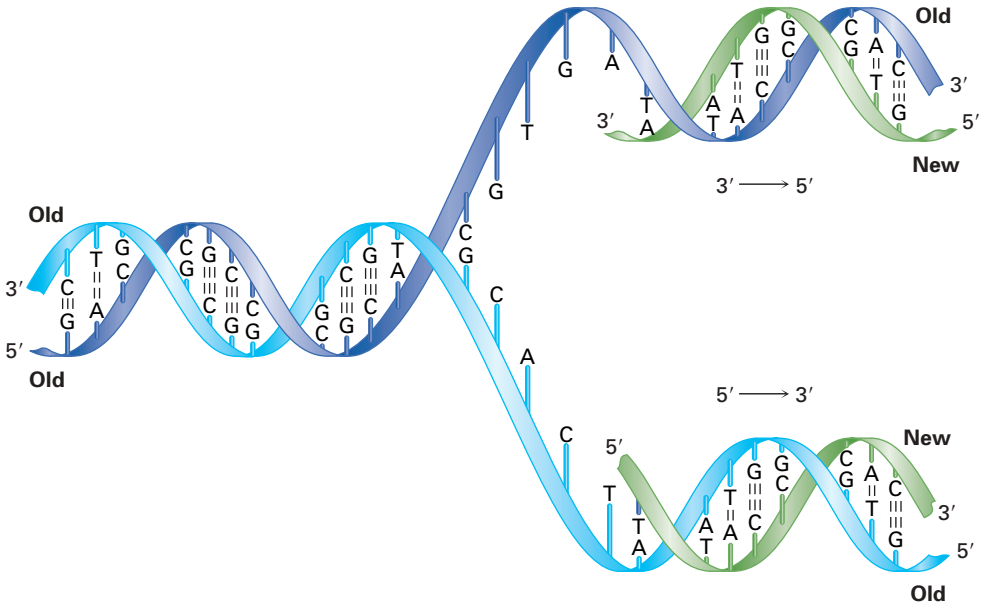
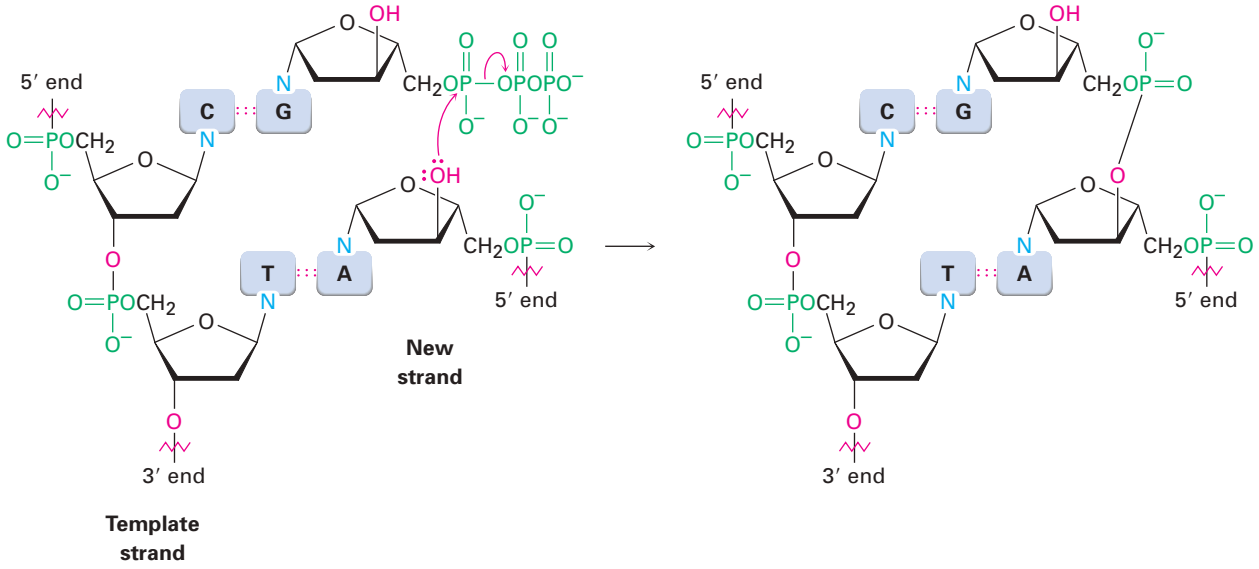


Figure 16.6 A representation of semiconservative DNA replication. The original double-stranded DNA partially unwinds, bases are exposed, nucleotides line up on each strand in a complementary manner, and two new strands begin to grow. Both strands are synthesized in the same 5′ → 3′ direction, one continuously and one in fragments.

Addition of nucleotides to a growing chain takes place in the 5′ → 3′ direction and is catalyzed by DNA polymerase. The key step is the addition of a

nucleoside 5′-triphosphate to the free 3′-hydroxyl group of the growing chain, with loss of a diphosphate leaving group.



The magnitude of the replication process is staggering. The nucleus of every human cell contains two copies of 22 chromosomes plus an additional 2 sex chromosomes, for a total of 46. Each chromosome consists of one very large DNA molecule, and the sum of the DNA in each of the two sets of chromosomes is estimated to be 3.0 billion base pairs, or 6.0 billion nucleotides. Despite the size of these enormous molecules, their base sequence is faithfully copied during replication. The entire copying process takes only a few hours and, after proofreading and repair, an error gets through only about once each 10 to 100 billion bases.

16.8 Transcription of DNA

As noted previously, RNA is structurally similar to DNA but contains ribose rather than deoxyribose and uracil rather than thymine. There are three major kinds of RNA, each of which serves a specific purpose. In addition, there are a number of small RNAs that appear to control a wide variety of important cellular functions. All RNA molecules are much smaller than DNA, and all remain single-stranded rather than double-stranded.

- **Messenger RNA (mRNA)** carries genetic messages from DNA to ribosomes, small particles in the cytoplasm of a cell where protein synthesis occurs.
- **Ribosomal RNA (rRNA)** complexed with protein provides the physical makeup of the ribosomes.
- **Transfer RNA (tRNA)** transports amino acids to the ribosomes, where they are joined together to make proteins.
- **Small RNAs**, also called *functional RNAs,* have a variety of functions within the cell, including silencing transcription and catalyzing chemical modifications of other RNA molecules.

The genetic information in DNA is contained in segments called *genes,* each of which consists of a specific nucleotide sequence that encodes a specific protein. The conversion of that information from DNA into proteins begins in the

nucleus of cells with the synthesis of mRNA by **transcription** of DNA. In bacteria, the process begins when RNA polymerase recognizes and binds to a *promoter sequence* on DNA, typically consisting of around 40 base pairs located upstream (5′) of the transcription start site. Within the promoter are two hexameric *consensus sequences,* one located 10 base pairs upstream of the start and the second located 35 base pairs upstream.

Following formation of the polymerase–promoter complex, several turns of the DNA double helix unwind, forming a bubble and exposing 14 or so base pairs of the two strands. Appropriate ribonucleotides then line up by hydrogen-bonding to their complementary bases on DNA, bond formation occurs in the 5′ → 3′ direction, the RNA polymerase moves along the DNA chain, and the growing RNA molecule unwinds from DNA (Figure 16.7). At any one time, about 12 base pairs of the growing RNA remain hydrogen-bonded to the DNA template.

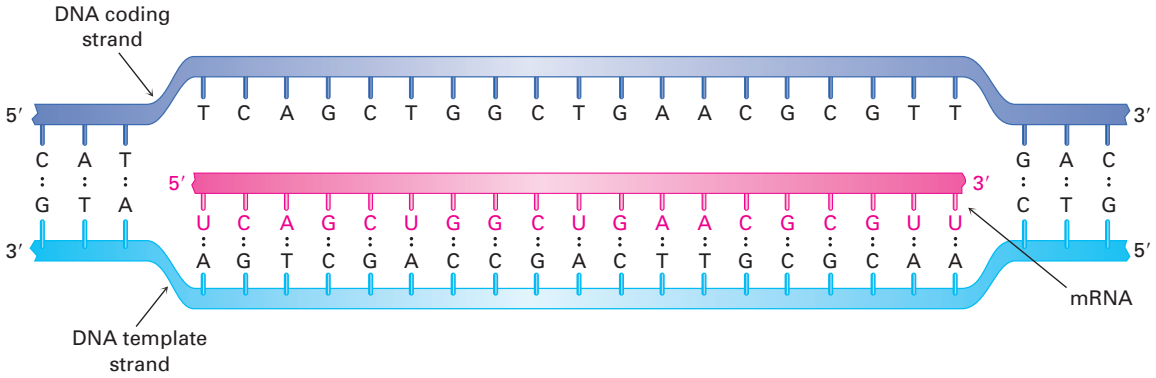


Figure 16.7 Synthesis of RNA using a DNA base segment as template.

Unlike what happens in DNA replication, where both strands are copied, only one of the two DNA strands is transcribed into mRNA. The DNA strand that contains the gene is often called the **sense strand**, or *coding strand,* and the DNA strand that gets transcribed to give RNA is called the **antisense strand**, or *noncoding strand.* Because the sense strand and the antisense strand in DNA are complementary, and because the DNA antisense strand and the newly formed RNA strand are also complementary, *the RNA molecule produced during transcription is a copy of the DNA sense strand.* That is, the complement of the complement is the same as the original. The only difference is that the RNA molecule has a U everywhere the DNA sense strand has a T.

Another part of the picture in vertebrates and flowering plants is that genes are often not continuous segments of the DNA chain. Instead, a gene will begin in one small section of DNA called an *exon,* then be interrupted by a noncoding section called an *intron,* and then take up again farther down the chain in another exon. The final mRNA molecule results only after the noncoded sections are cut out of the transcribed mRNA and the remaining pieces are joined together by spliceosomes. The gene for triose phosphate isomerase in maize, for instance, contains eight noncoding introns accounting for approximately 70% of the DNA base pairs and nine coding exons accounting for only 30% of the base pairs.

Worked Example 16.4 RNA/DNA Complementary Sequences

What RNA base sequence is complementary to the following DNA base sequence?

(5′) TAAGCCGTG (3′)

Strategy Go through the sequence replacing A by U, G by C, T by A, and C by G.

Solution

Original DNA: (5′) TAAGCCGTG (3′)
Complementary RNA: (3′) AUUCGGCAC (5′)

Problem 16.11 Show how uracil can form strong hydrogen bonds to adenine, just as thymine can.

Problem 16.12 What RNA base sequence is complementary to the following DNA base sequence?

(5′) GATTACCGTA (3′)

Problem 16.13 From what DNA base sequence was the following RNA sequence transcribed?

(5′) UUCGCAGAGU (3′)

16.9 Translation of RNA: Protein Biosynthesis

The primary cellular function of mRNA is to direct biosynthesis of the thousands of diverse peptides and proteins required by an organism—as many as 500,000 in a human. The mechanics of protein biosynthesis take place on ribosomes, small granular particles in the cytoplasm of a cell that consist of about 60% ribosomal RNA and 40% protein.

The specific ribonucleotide sequence in mRNA forms a message that determines the order in which amino acid residues are to be joined. Each "word," or **codon**, along the mRNA chain consists of a sequence of three ribonucleotides that is specific for a given amino acid. For example, the series UUC on mRNA is a codon directing incorporation of the amino acid phenylalanine into the growing protein. Of the $4^3 = 64$ possible triplets of the four bases in RNA, 61 code for specific amino acids and 3 code for chain termination. Table 16.3 shows the meaning of each codon.

Table 16.3 Codon Assignments of Base Triplets

First base (5′ end)	Second base	Third base (3′ end) U	C	A	G
U	U	Phe	Phe	Leu	Leu
	C	Ser	Ser	Ser	Ser
	A	Tyr	Tyr	Stop	Stop
	G	Cys	Cys	Stop	Trp
C	U	Leu	Leu	Leu	Leu
	C	Pro	Pro	Pro	Pro
	A	His	His	Gln	Gln
	G	Arg	Arg	Arg	Arg
A	U	Ile	Ile	Ile	Met
	C	Thr	Thr	Thr	Thr
	A	Asn	Asn	Lys	Lys
	G	Ser	Ser	Arg	Arg
G	U	Val	Val	Val	Val
	C	Ala	Ala	Ala	Ala
	A	Asp	Asp	Glu	Glu
	G	Gly	Gly	Gly	Gly

The message embedded in mRNA is read by transfer RNA (tRNA) in a process called **translation**. There are 61 different tRNAs, one for each of the 61 codons that specifies an amino acid. A typical tRNA is single-stranded and has roughly the shape of a cloverleaf, as shown in Figure 16.8. It consists of about 70 to 100 ribonucleotides and is bonded to a specific amino acid by an ester linkage through the 3′ hydroxyl on ribose at the 3′ end of the tRNA. Each tRNA also contains on its middle leaf a segment called an **anticodon**, a sequence of three ribonucleotides complementary to the codon sequence. For example, the codon sequence UUC present on mRNA is read by a phenylalanine-bearing tRNA having the complementary anticodon base sequence GAA. [Remember that nucleotide sequences are written in the 5′ → 3′ direction, so the sequence in an anticodon must be reversed. That is, the complement to (5′) UUC (3′) is (3′) AAG (5′), which is written as (5′) GAA (3′).]

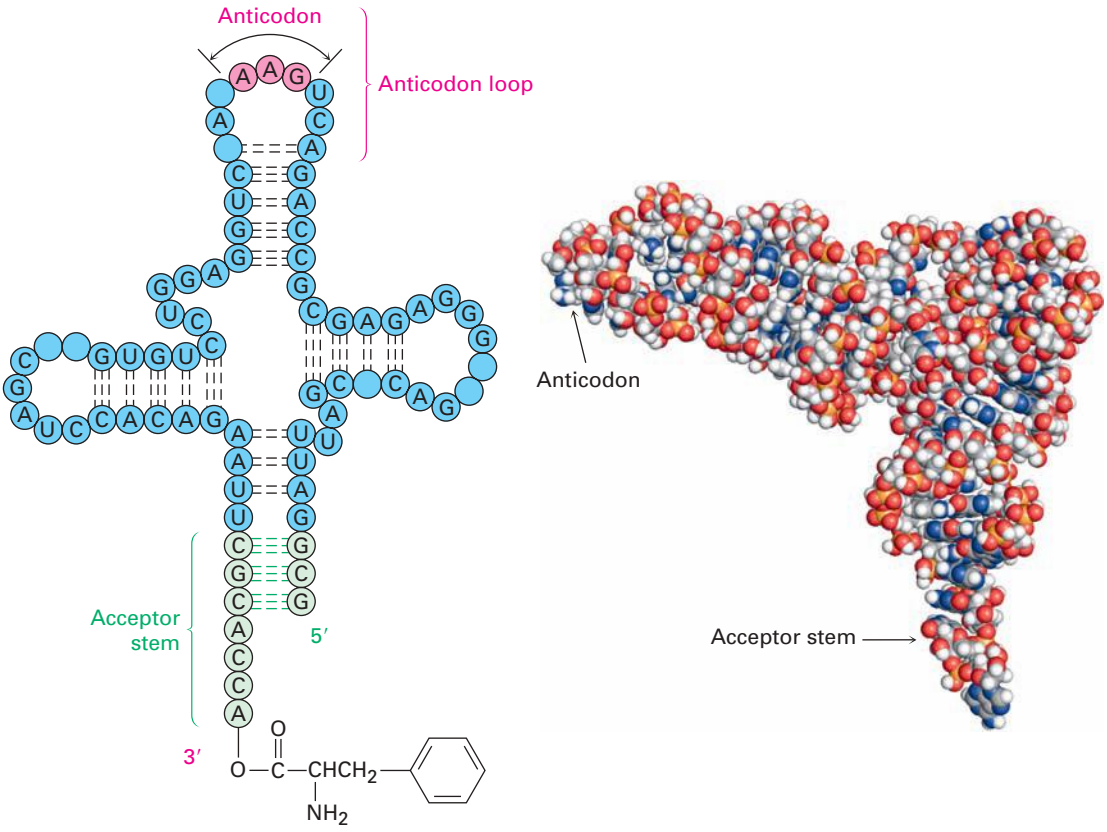


Figure 16.8 Structure of a tRNA molecule. The tRNA molecule is roughly cloverleaf-shaped and contains an anticodon triplet on one "leaf" and an amino acid unit attached covalently at its 3′ end. The example shown is a yeast tRNA that codes for phenylalanine. The nucleotides not specifically identified are chemically modified analogs of the four common ribonucleotides.

As each successive codon on mRNA is read, different tRNAs bring the correct amino acids into position for enzyme-mediated transfer to the growing peptide. When synthesis of the proper protein is completed, a "stop" codon signals the end and the protein is released from the ribosome. The process is illustrated in Figure 16.9.

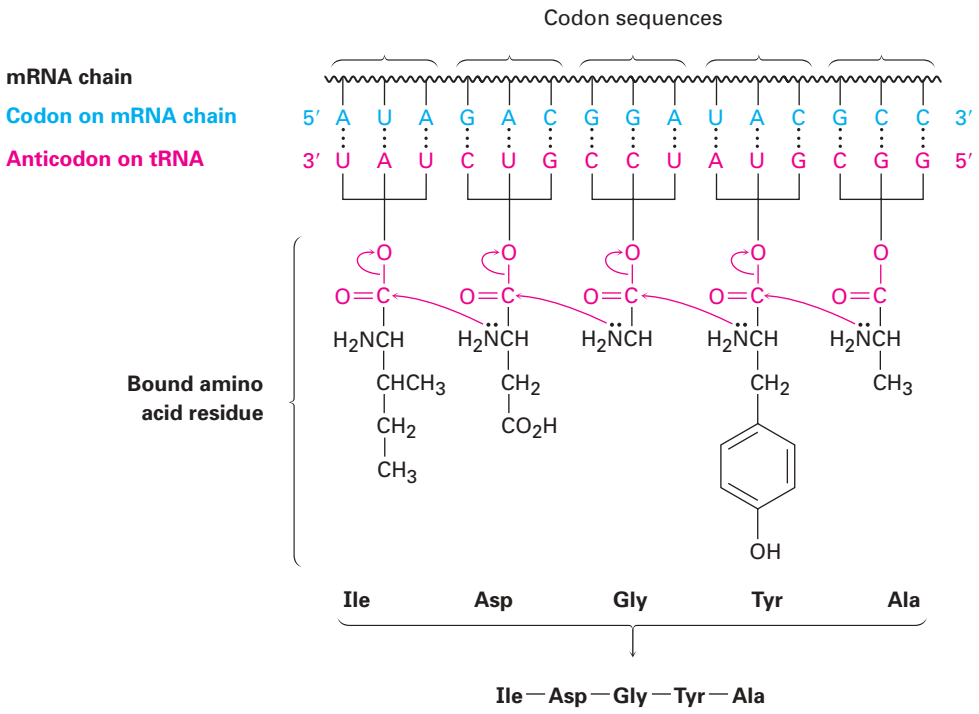


Figure 16.9 A representation of protein biosynthesis. The codon base sequences on mRNA are read by tRNAs containing complementary anticodon base sequences. Transfer RNAs assemble the proper amino acids into position for incorporation into the growing peptide.

Worked Example 16.5 Codon Sequences for Amino Acids

Give a codon sequence for valine.

Solution According to Table 16.3, there are four codons for valine: GUU, GUC, GUA, and GUG.

Worked Example 16.6 Finding the Amino Acid Sequence Transcribed from DNA

What amino acid sequence is coded by the following segment of a DNA sense strand?

(5′) CTA-ACT-AGC-GGG-TCG-CCG (3′)

Strategy The mRNA produced during translation is a copy of the DNA sense strand, with each T replaced by U. Thus, the mRNA has the sequence

(5′) CUA-ACU-AGC-GGG-UCG-CCG (3′)

Each set of three bases forms a codon, whose meaning can be found in Table 16.3.

Solution Leu-Thr-Ser-Gly-Ser-Pro

Problem 16.14 List codon sequences for the following amino acids:
(a) Ala **(b)** Phe **(c)** Leu **(d)** Tyr

Problem 16.15 What amino acid sequence is coded by the following mRNA base sequence?

(5′) CUU-AUG-GCU-UGG-CCC-UAA (3′)

Problem 16.16 What anticodon sequences of tRNAs are coded by the mRNA in Problem 16.15?

Problem 16.17 What was the base sequence in the original DNA strand on which the mRNA sequence in Problem 16.15 was made?

16.10 DNA Sequencing

One of the greatest scientific revolutions in history is now underway in molecular biology, as scientists are learning how to manipulate and harness the genetic machinery of organisms. None of the extraordinary advances of the past two decades would have been possible, however, were it not for the discovery in 1977 of methods for sequencing immense DNA chains.

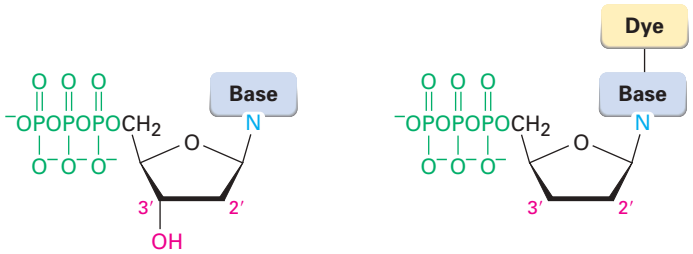
The first step in DNA sequencing is to cleave the enormous chain at known points to produce smaller, more manageable pieces, a task accomplished by the use of *restriction endonucleases.* Each different restriction enzyme, of which more than 3500 are known and approximately 200 are commercially available, cleaves a DNA molecule at a point in the chain where a specific base sequence occurs. For example, the restriction enzyme *Alu*I cleaves between G and C in the four-base sequence AG-CT. Note that the sequence is a *palindrome,* meaning that the *sequence* (5′)-AGCT-(3′) is the same as its *complement* (3′)-TCGA-(5′) when both are read in the same 5′ → 3′ direction. The same is true for other restriction endonucleases.

If the original DNA molecule is cut with another restriction enzyme that has a different specificity for cleavage, still other segments are produced whose sequences partially overlap those produced by the first enzyme. Sequencing of all the segments, followed by identification of the overlapping regions, allows complete DNA sequencing.

Two methods of DNA sequencing are available. The *Maxam–Gilbert method* uses chemical techniques, while the **Sanger dideoxy method** uses enzymatic reactions. The Sanger method is the more commonly used of the two and is the method responsible for sequencing the entire human genome of 3.0 billion base pairs. In commercial sequencing instruments, the dideoxy method begins with a mixture of the following:

- The restriction fragment to be sequenced
- A small piece of DNA called a *primer,* whose sequence is complementary to that on the 3′ end of the restriction fragment
- The four 2′-deoxyribonucleoside triphosphates (dNTPs)
- Very small amounts of the four 2′,3′-*dideoxy*ribonucleoside triphosphates (ddNTPs), each of which is labeled with a fluorescent dye of a

different color. (A 2′,3′-dideoxyribonucleoside triphosphate is one in which both 2′ and 3′ –OH groups are missing from ribose.)



A 2′-deoxyribonucleoside triphosphate (dNTP)

A 2′,3′-dideoxyribonucleoside triphosphate (ddNTP)

DNA polymerase is added to the mixture, and a strand of DNA complementary to the restriction fragment begins to grow from the end of the primer. Most of the time, only normal deoxyribonucleotides are incorporated into the growing chain because of their much higher concentration in the mixture, but every so often, a dideoxyribonucleotide is incorporated. When that happens, DNA synthesis stops because the chain end no longer has a 3′-hydroxyl group for adding further nucleotides.

When reaction is complete, the product consists of a mixture of DNA fragments of all possible lengths, each terminated by one of the four dye-labeled dideoxyribonucleotides. This product mixture is then separated according to the size of the pieces by gel electrophoresis (Section 15.2), and the identity of the terminal dideoxyribonucleotide in each piece—and thus the sequence of the restriction fragment—is determined by noting the color with which it fluoresces. Figure 16.10 shows a typical result.

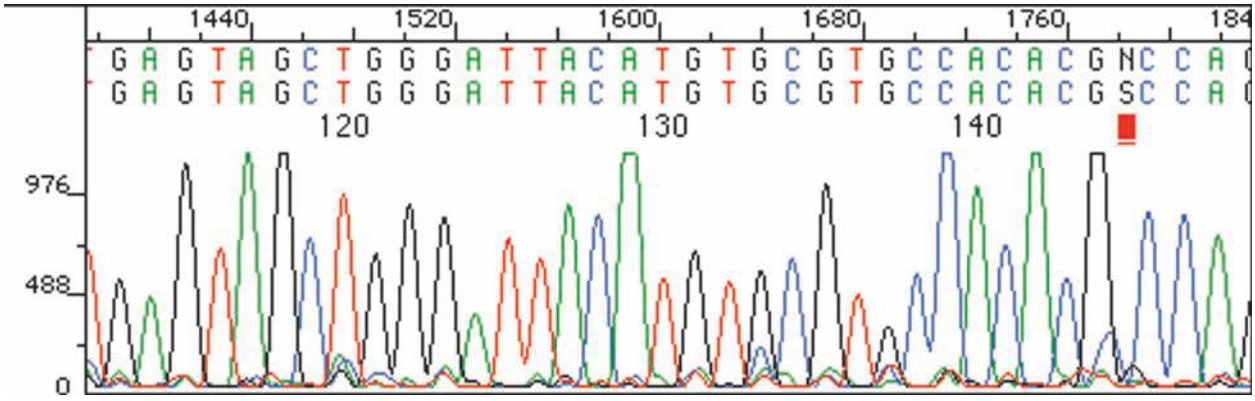


Figure 16.10 The sequence of a restriction fragment determined by the Sanger dideoxy method can be read simply by noting the colors of the dye attached to each of the various terminal nucleotides.

So efficient is the automated dideoxy method that sequences up to 1100 nucleotides in length, with a throughput of up to 19,000 bases per hour, can be sequenced with 98% accuracy. After a decade of work, preliminary sequence information for

the entire human genome of 3.0 billion base pairs was announced early in 2001 and complete information was released in 2003. More recently, the genome sequencing of specific individuals, including that of James Watson, discoverer of the double helix, has been accomplished.

Remarkably, our genome appears to contain only about 21,000 genes, less than one-fourth the previously predicted number and only about twice the number found in the common roundworm. It's also interesting to note that the number of genes in a human (21,000) is much smaller than the number of kinds of proteins (perhaps 500,000). The discrepancy arises because most proteins are modified in various ways after translation (posttranslational modifications), so a single gene can ultimately give many different proteins.

16.11 The Polymerase Chain Reaction

It often happens that only a tiny amount of DNA can be obtained directly, as might occur at a crime scene, so methods for obtaining larger amounts are sometimes needed to carry out the sequencing and characterization. The invention of the **polymerase chain reaction (PCR)** in 1986 has been described as being to genes what Gutenberg's invention of the printing press was to the written word. Just as the printing press produces multiple copies of a book, PCR produces multiple copies of a given DNA sequence. Starting from less than 1 *picogram* of DNA with a chain length of 10,000 nucleotides (1 pg = 10^{-12} g; about 10^5 molecules), PCR makes it possible to obtain several micrograms (1 μg = 10^{-6} g; about 10^{11} molecules) in just a few hours.

The key to the polymerase chain reaction is *Taq* DNA polymerase, a heat-stable enzyme isolated from the thermophilic bacterium *Thermus aquaticus* found in a hot spring in Yellowstone National Park. *Taq* polymerase is able to take a single strand of DNA that has a short, primer segment of complementary chain at one end and then finish constructing the entire complementary strand. The overall process takes three steps, as shown in Figure 16.11. (More recently, improved heat-stable DNA polymerase enzymes have become available, including Vent polymerase and *Pfu* polymerase, both isolated from bacteria growing near geothermal vents in the ocean floor. The error rate of both enzymes is substantially less than that of *Taq*.)

STEP 1 The double-stranded DNA to be amplified is heated in the presence of *Taq* polymerase, Mg^{2+} ion, the four deoxynucleotide triphosphate monomers (dNTPs), and a large excess of two short oligonucleotide primers of about 20 bases each. Each primer is complementary to the sequence at the end of one of the target DNA segments. At 95 °C, double-stranded DNA denatures, spontaneously breaking apart into two single strands.

STEP 2 The temperature is lowered to between 37 and 50 °C, allowing the primers, because of their relatively high concentration, to anneal by hydrogen-bonding to their complementary sequence at the end of each target strand.

STEP 3 The temperature is then raised to 72 °C, and *Taq* polymerase catalyzes the addition of further nucleotides to the two primed DNA strands. When replication of each strand is finished, *two* copies of the original DNA now exist. Repeating the denature–anneal–synthesize cycle a second time yields four DNA copies, repeating a third time yields eight copies, and so on, in an exponential series.

PCR has been automated, and 30 or so cycles can be carried out in an hour, resulting in a theoretical amplification factor of 2^{30} ($\sim 10^9$). In practice, however, the efficiency of each cycle is less than 100%, and an experimental amplification of about 10^6 to 10^8 is routinely achieved for 30 cycles.

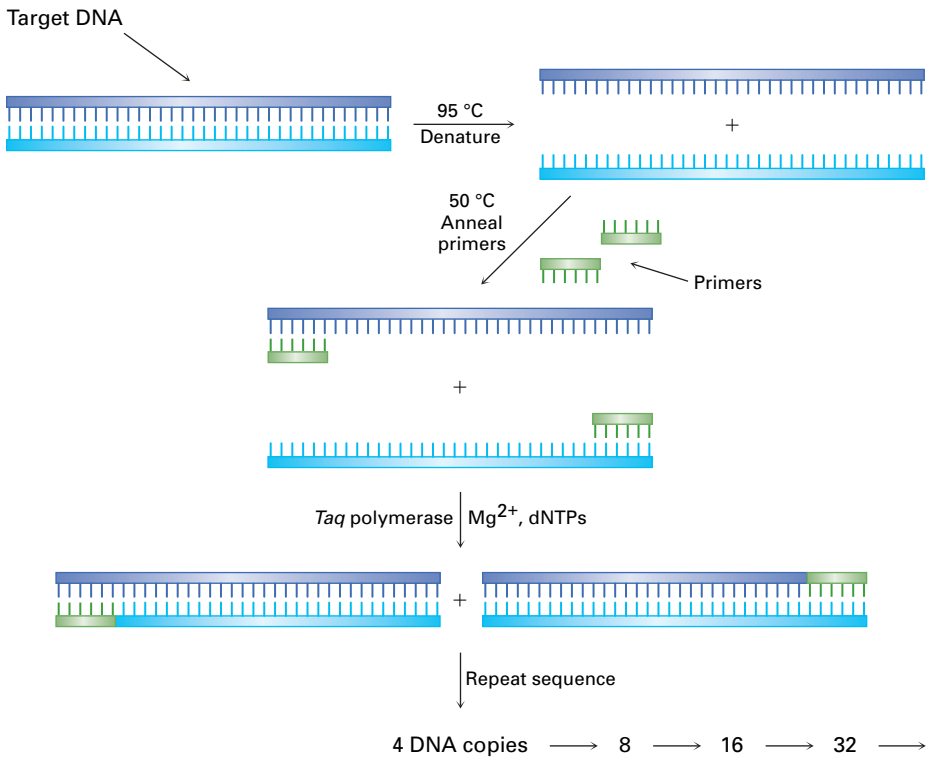


Figure 16.11 The polymerase chain reaction. Details are explained in the text.

INTERLUDE

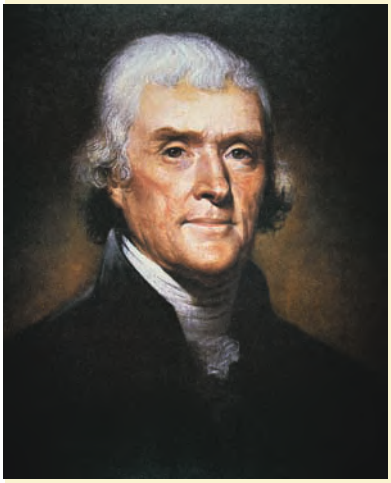
DNA Fingerprinting

The invention of DNA sequencing has affected society in many ways, few more dramatic than those stemming from the development of *DNA fingerprinting*. DNA fingerprinting arose from the discovery in 1984 that human genes contain short, repeating sequences of noncoding DNA, called *short tandem repeat* (STR) loci. Furthermore, the STR loci are slightly different for every individual, except identical twins. By sequencing these loci, a pattern unique to each person can be obtained.

Perhaps the most common and well-publicized use of DNA fingerprinting is that carried out by crime laboratories to link suspects to biological evidence—blood, hair follicles, skin, or semen—found at a crime scene. Thousands of court cases have now been decided based on DNA evidence.

continued

INTERLUDE



Rembrandt Peale/Getty Images

Historians have wondered for many years whether Thomas Jefferson fathered a child by Sally Hemings. DNA fingerprinting evidence obtained in 1998 suggests that he did.

For use in criminal cases, forensic laboratories in the United States have agreed on 13 core STR loci that are most accurate for identification of an individual. Based on these 13 loci, a Combined DNA Index System (CODIS) has been established to serve as a registry of convicted offenders. When a DNA sample is obtained from a crime scene, the sample is subjected to cleavage with restriction endonucleases to cut out fragments containing the STR loci, the fragments are amplified using the polymerase chain reaction, and the sequences of the fragments are determined.

If the profile of sequences from a known individual and the profile from DNA obtained at a crime scene match, the probability is approximately 82 billion to 1 that the DNA is from the same individual. In paternity cases, where the DNA of father and offspring are related but not fully identical, the identity of the father can be established with a probability of around 100,000 to 1. Even after several generations have passed, paternity can still be implied by DNA analysis of the Y chromosome of direct male-line descendants. The most well-known such case is that of Thomas Jefferson, who likely fathered a child by his slave Sally Hemings. Although Jefferson himself has no male-line descendants, DNA analysis of the male-line descendants of Jefferson's paternal uncle contained the same Y chromosome as a male-line descendant of Eston Hemings, the youngest son of Sally Hemings. Thus, a mixing of the two genomes is clear, although the male individual responsible for that mixing can't be conclusively identified.

Among its many other applications, DNA fingerprinting is widely used for the diagnosis of genetic disorders, both prenatally and in newborns. Cystic fibrosis, hemophilia, Huntington's disease, Tay–Sachs disease, sickle cell anemia, and thalassemia are among the many diseases that can be detected, enabling early treatment of an affected child. Furthermore, by studying the DNA fingerprints of relatives with a history of a particular disorder, it's possible to identify DNA patterns associated with the disease and perhaps obtain clues for eventual cure. In addition, the U.S. Department of Defense now requires blood and saliva samples from all military personnel. The samples are stored, and DNA is extracted if the need for identification of a casualty arises.

Summary and Key Words

In this chapter, we've looked at lipids and nucleic acids, completing our coverage of the four main classes of biological molecules. **Lipids** are the naturally occurring substances isolated from plants and animals by extraction with organic solvents. Animal fats and vegetable oils are the most widely occurring lipids. Both are **triacylglycerols**—triesters of glycerol with long-chain **fatty acids**. **Phospholipids** are esters of phosphoric acid and have either glycerol or an amino alcohol for their backbone.

Steroids are plant and animal lipids with a characteristic tetracyclic skeleton. Steroids occur widely in body tissue and have many different kinds of physiological properties. Among the more important kinds of

nucleic acid 548
nucleoside 549
nucleotide 549
phospholipid 544
polymerase chain reaction (PCR) 562
polyunsaturated fatty acid 541
replication 554
ribonucleic acid (RNA) 548
ribosomal RNA (rRNA) 555
Sanger dideoxy method 560
sense strand 556
small RNAs 555
steroid 546
transcription 556
transfer RNA (tRNA) 555
translation 558
triacylglycerol 539
wax 539

steroids are the sex hormones (*androgens* and *estrogens*) and the adrenocortical hormones.

The **nucleic acids**, **DNA** (**deoxyribonucleic acid**) and **RNA** (**ribonucleic acid**), are biological polymers that act as chemical carriers of an organism's genetic information. Enzyme-catalyzed hydrolysis of nucleic acids yields **nucleotides**, the monomer units from which RNA and DNA are constructed. Further enzyme-catalyzed hydrolysis of the nucleotides yields **nucleosides** plus phosphate. Nucleosides, in turn, consist of a purine or pyrimidine base linked to C1 of an aldopentose sugar—ribose in RNA and 2-deoxyribose in DNA. The nucleotides are joined by phosphodiester bonds between the 5′ phosphate of one nucleotide and the 3′ hydroxyl on the sugar of another nucleotide.

Molecules of DNA consist of two complementary strands held together by hydrogen bonds between heterocyclic bases on the different strands and coiled into a **double helix**. Adenine (A) and thymine (T) form hydrogen bonds to each other, as do cytosine (C) and guanine (G).

Three processes take place in deciphering the genetic information of DNA:

- **Replication** of DNA is the process by which identical DNA copies are made. The DNA double helix unwinds, complementary deoxyribonucleotides line up in order, and two new DNA molecules are produced.
- **Transcription** is the process by which RNA is produced to carry genetic information from the nucleus to the ribosomes. A short segment of the DNA double helix unwinds, and complementary ribonucleotides line up to produce **messenger RNA (mRNA)**.
- **Translation** is the process by which mRNA directs protein synthesis. Each mRNA is divided into **codons**, ribonucleotide triplets that are recognized by small amino acid carrying molecules of **transfer RNA (tRNA)**, which deliver the appropriate amino acids needed for protein synthesis.

Sequencing of DNA is carried out by the **Sanger dideoxy method**. Small amounts of DNA can be amplified by a factor of 10^6 using the **polymerase chain reaction (PCR)**.

Exercises

Visualizing Chemistry

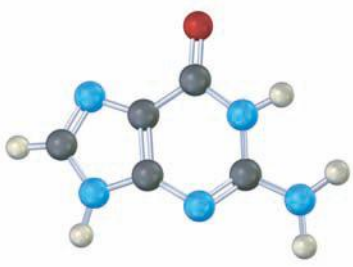
(Problems 16.1–16.17 appear within the chapter.)

OWL

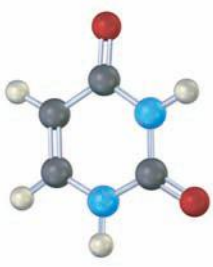
Interactive versions of these problems are assignable in OWL.

16.18 Identify the following bases, and tell whether each is found in DNA, RNA, or both:

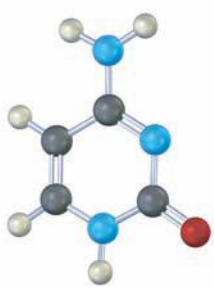
(a)



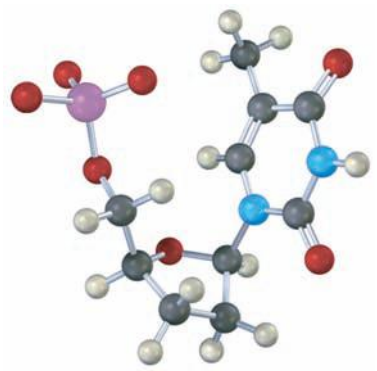
(b)



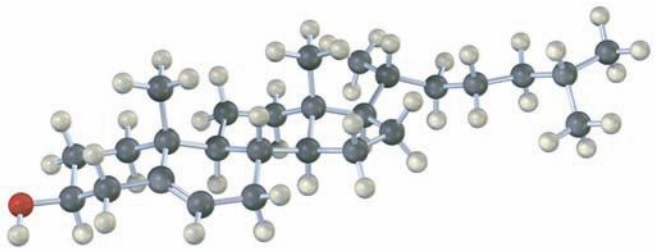
(c)



16.19 Identify the following nucleotide, and tell how it is used:



16.20 Cholesterol has the following structure. Tell whether the –OH group is axial or equatorial.



Additional Problems

Lipids

16.21 Write representative structures for the following:
(a) A fat **(b)** A vegetable oil **(c)** A steroid

16.22 Write the structures of the following molecules:
(a) Sodium stearate
(b) Ethyl linoleate
(c) Glyceryl dioleopalmitate

16.23 Show the products you would expect to obtain from the reaction of glyceryl trioleate with the following:
(a) Excess Br_2 in CCl_4
(b) H_2/Pd
(c) NaOH, H_2O
(d) $KMnO_4$, H_3O^+
(e) $LiAlH_4$, then H_3O^+

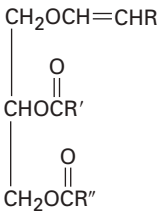
16.24 How would you convert oleic acid into the following substances?
(a) Methyl oleate **(b)** Methyl stearate **(c)** Nonanedioic acid

16.25 Eleostearic acid, $C_{18}H_{30}O_2$, is a rare fatty acid found in tung oil. On oxidation with $KMnO_4$, eleostearic acid yields 1 part pentanoic acid, 2 parts oxalic acid ($HO_2C—CO_2H$), and 1 part nonanedioic acid. Propose a structure for eleostearic acid.

16.26 Stearolic acid, $C_{18}H_{32}O_2$, yields oleic acid on catalytic hydrogenation over the Lindlar catalyst. Propose a structure for stearolic acid.

16.27 Spermaceti, a fragrant substance from sperm whales, was much used in cosmetics until it was banned in 1976 to protect whales from extinction. Chemically, spermaceti is cetyl palmitate, the ester of cetyl alcohol ($C_{16}H_{33}OH$) with palmitic acid. Draw its structure.

16.28 The plasmalogens are a group of lipids found in nerve and muscle cells. How do plasmalogens differ from fats?

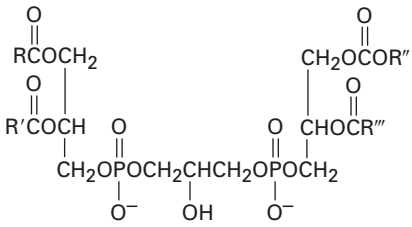


A plasmalogen

16.29 Draw the products you would obtain from treatment of cholesterol with the following reagents:
(a) Br_2 **(b)** H_2, Pd catalyst **(c)** CH_3COCl, pyridine

16.30 If the average molecular weight of soybean oil is 1500, how many grams of NaOH are needed to saponify 5.00 g of the oil?

16.31 Cardiolipins are a group of lipids found in heart muscles. What products would be formed if all ester bonds, including phosphates, were saponified by treatment with aqueous NaOH?



A cardiolipin

Nucleic Acids

16.32 The DNA from sea urchins contains about 32% A and about 18% G. What percentages of T and C would you expect in sea urchin DNA? Explain.

16.33 What DNA sequence is complementary to the following sequence?

(5′) GAAGTTCATGC (3′)

16.34 Give codons for the following amino acids:
(a) Ile **(b)** Asp **(c)** Thr

16.35 Draw the complete structure of the ribonucleotide codon UAC. For what amino acid does this sequence code?

16.36 Draw the complete structure of the deoxyribonucleotide sequence from which the mRNA codon in Problem 16.35 was transcribed.

16.37 What amino acids do the following ribonucleotide codons code for?
(a) AAU **(b)** GAG **(c)** UCC **(d)** CAU **(e)** ACC

16.38 From what DNA sequences were each of the mRNA codons in Problem 16.37 transcribed?

16.39 What anticodon sequences of tRNAs are coded by each of the codons in Problem 16.37?

16.40 The codon UAA stops protein synthesis. Why does the sequence UAA in the following stretch of mRNA not cause any problems?

-GCA-UUC-GAG-GUA-ACG-CCC-

16.41 If the gene sequence -TAA-CCG-GAT- on DNA were miscopied during replication and became -TGA-CCG-GAT-, what effect would the mutation have on the sequence of the protein produced?

16.42 Give an mRNA sequence that codes for synthesis of metenkephalin, a small peptide with morphine-like properties:

Tyr-Gly-Gly-Phe-Met

16.43 Give a DNA gene sequence (sense strand) that will code for metenkephalin (Problem 16.42).

16.44 Human and horse insulin both have two polypeptide chains with one chain containing 21 amino acids and the other containing 30 amino acids. How many nitrogen bases are present in the DNA to code for each chain?

16.45 Human and horse insulin (see Problem 16.44) differ in primary structure at two amino acids: at the 9th position in one chain (human has Ser and horse has Gly) and at the 30th position in the other chain (human has Thr and horse has Ala). How must the DNA differ?

General Problems

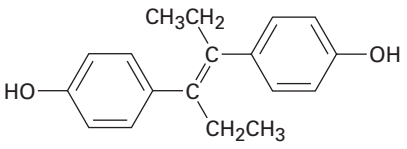
16.46 What amino acid sequence is coded by the following mRNA sequence?

CUA-GAC-CGU-UCC-AAG-UGA

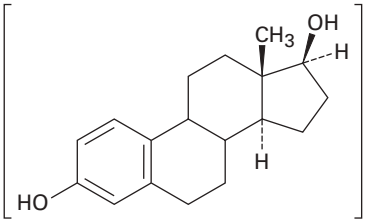
16.47 What anticodon sequences of tRNAs are coded by the mRNA in Problem 16.46? What was the base sequence in the original DNA strand on which this mRNA was made? What was the base sequence in the DNA strand complementary to that from which this mRNA was made?

16.48 Look up the structure of angiotensin II in Worked Example 15.4 on page 516, and give an mRNA sequence that codes for its synthesis.

16.49 Diethylstilbestrol (DES) exhibits estradiol-like activity even though it is structurally unrelated to steroids. Once used widely as an additive in animal feed, DES has been implicated as a causative agent in several types of cancers. Show how DES can be drawn so that it is sterically similar to estradiol.



Diethylstilbestrol

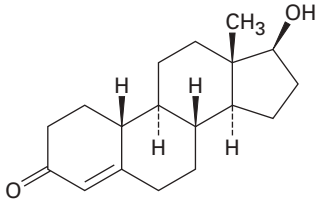


Estradiol

16.50 How many chirality centers are present in estradiol (see Problem 16.49)? Label them.

16.51 What products would you obtain from reaction of estradiol (Problem 16.49) with the following reagents?
(a) NaOH, then CH_3I **(b)** CH_3COCl, pyridine **(c)** Br_2 (1 equiv)

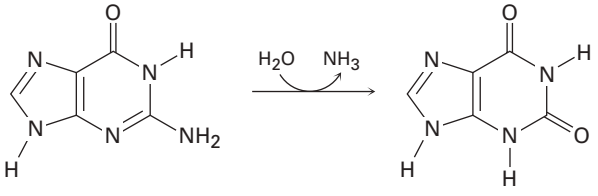
16.52 Nandrolone is an anabolic steroid sometimes taken by athletes to build muscle mass. Compare the structures of nandrolone and testosterone (page 547), and point out their structural similarities.



Nandrolone
(an anabolic steroid)

16.53 Draw the structure of cyclic adenosine monophosphate (cAMP), a messenger involved in the regulation of glucose production in the body. Cyclic AMP has a phosphate ring connecting the 3′ and 5′ hydroxyl groups on adenosine.

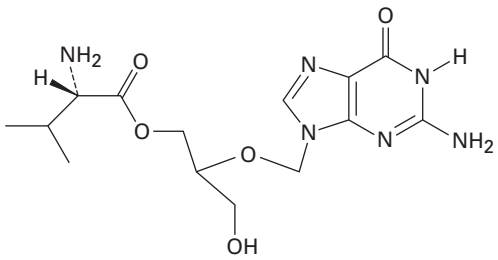
16.54 One of the steps in the metabolic degradation of guanine is hydrolysis to give xanthine. Propose a mechanism.



Guanine **Xanthine**

IN THE MEDICINE CABINET

16.55 Valganciclovir is an antiviral agent used for the treatment of cytomegalovirus. Called a *prodrug,* valganciclovir is inactive by itself but is rapidly converted in the intestine by hydrolysis of its ester bond to produce an active drug called ganciclovir, along with an amino acid.



Valganciclovir

(a) What amino acid is produced by hydrolysis of the ester bond in valganciclovir?
(b) What is the structure of ganciclovir?
(c) What atoms present in the nucleotide deoxyguanine are missing from ganciclovir?
(d) What role do the atoms missing from deoxyguanine play in DNA replication?
(e) How might valganciclovir interfere with DNA synthesis?

IN THE FIELD

16.56 Weeds derive resistance to the herbicide atrazine by a mutation in which amino acid substitution of a serine residue by a glycine occurs.
(a) Identify the six codons that correspond to serine and the four codons that correspond to glycine.
(b) What is the minimum number of base mutations of DNA required for the evolution of resistance?