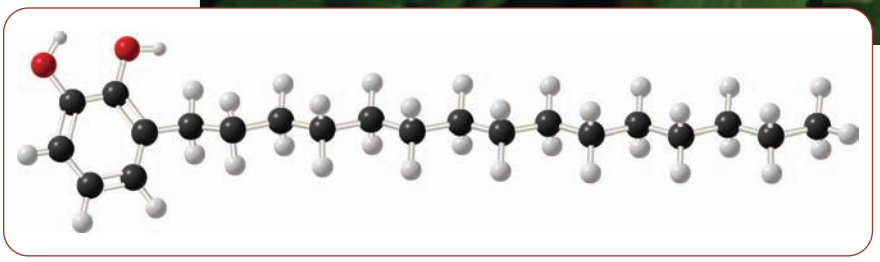
21

Phenols and Aryl Halides

Nucleophilic Aromatic Substitution

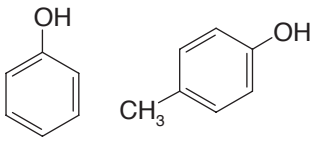


A phenol that is a blistering agent in poison ivy.

In this chapter we shall study phenols and aryl halides. A phenol contains a hydroxyl group directly bonded to a benzene ring. An aryl halide contains a halogen directly bonded to a benzene ring. As we learn about these classes of compounds we shall learn some new reactions, including nucleophilic aromatic substitution and the Claisen rearrangement, and have opportunities to review reactions that we have studied previously. We shall also see that phenols have widely varying roles in nature, from hormones and antibiotics to the blistering agents of poison ivy, like the molecule shown above. Aryl halides also have important properties, although some of them, such as polychlorinated and polybrominated biphenyls, have proved to have harmful effects on the environment. We begin with consideration of phenols.

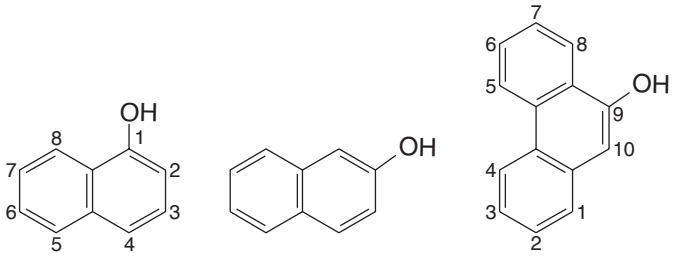
21.1 Structure and Nomenclature of Phenols

Compounds that have a hydroxyl group directly attached to a benzene ring are called **phenols.** Thus, **phenol** is the specific name for hydroxybenzene, and it is the general name for the family of compounds derived from hydroxybenzene:



Phenol **4-Methylphenol (a phenol)**

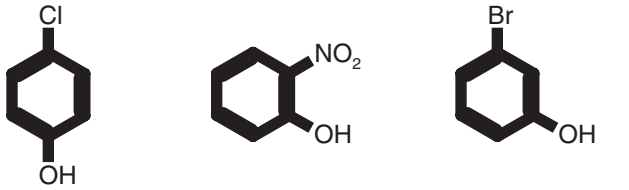
Compounds that have a hydroxyl group attached to a polycyclic benzenoid ring are chemically similar to phenols, but they are called **naphthols** and **phenanthrols**, for example:



1-Naphthol (α-naphthol) **2-Naphthol (β-naphthol)** **9-Phenanthrol**

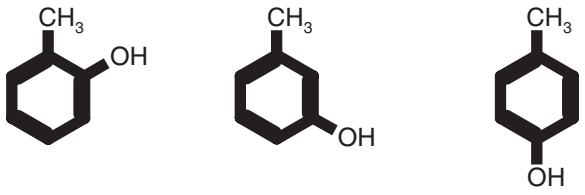
21.1A Nomenclature of Phenols

We studied the nomenclature of some phenols in Chapter 14. In many compounds *phenol* is the base name:



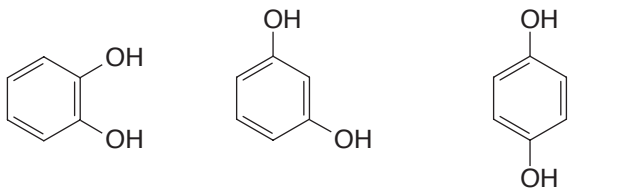
4-Chlorophenol (*p*-chlorophenol) **2-Nitrophenol (*o*-nitrophenol)** **3-Bromophenol (*m*-bromophenol)**

The methylphenols are commonly called *cresols*:



2-Methylphenol (*o*-cresol) **3-Methylphenol (*m*-cresol)** **4-Methylphenol (*p*-cresol)**

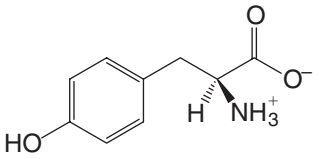
The benzenediols also have common names:



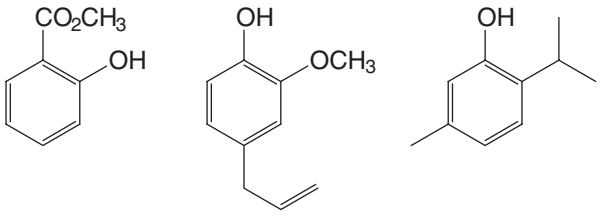
1,2-Benzenediol (catechol) **1,3-Benzenediol (resorcinol)** **1,4-Benzenediol (hydroquinone)**

21.2 Naturally Occurring Phenols

Phenols and related compounds occur widely in nature. Tyrosine is an amino acid that occurs in proteins. (See "The Chemistry of . . . Iodine Incorporation in Thyroxine Biosynthesis" in Chapter 15.) Methyl salicylate is found in oil of wintergreen, eugenol is found in oil of cloves, and thymol is found in thyme.

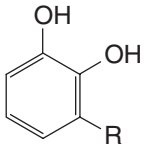


L-Tyrosine



Methyl salicylate (oil of wintergreen) **Eugenol (oil of cloves)** **Thymol (thyme)**

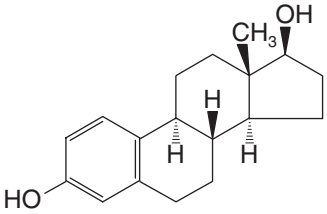
The urushiols are blistering agents (vesicants) found in poison ivy.



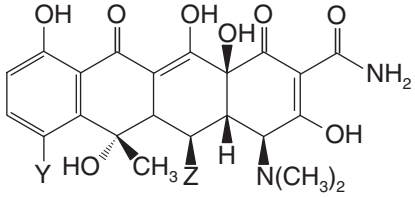
Urushiols

R = $-(CH_2)_{14}CH_3$,
$-(CH_2)_7CH{=}CH(CH_2)_5CH_3$, or
$-(CH_2)_7CH{=}CHCH_2CH{=}CH(CH_2)_2CH_3$, or
$-(CH_2)_7CH{=}CHCH_2CH{=}CHCH{=}CHCH_3$ or
$-(CH_2)_7CH{=}CHCH_2CH{=}CHCH_2CH{=}CH_2$

Estradiol is a female sex hormone, and the tetracyclines are important antibiotics.



Estradiol



Tetracyclines
(Y = Cl, Z = H; Aureomycin)
(Y = H, Z = OH; Terramycin)

21.3 Physical Properties of Phenols

The presence of hydroxyl groups in phenols means that phenols are like alcohols (Section 11.2) in some respects. For example, they are able to form strong intermolecular hydrogen bonds, and therefore have higher boiling points than hydrocarbons of the same molecular weight. Phenol (bp 182°C) has a boiling point more than 70°C higher than toluene (bp 110.6°C), even though the two compounds have almost the same molecular weight. Phenols are also modestly soluble in water because of their ability to form strong hydrogen bonds with water molecules.

21.4 Synthesis of Phenols

21.4A Laboratory Synthesis

The most important laboratory synthesis of phenols is by hydrolysis of arenediazonium salts (Section 20.7E). This method is highly versatile, and the conditions required for the diazotization step and the hydrolysis step are mild. This means that other groups present on the ring are unlikely to be affected.

General Reaction

$$\text{Ar—NH}_2 \xrightarrow{\text{HONO}} \text{Ar—}\overset{+}{\text{N}}\text{≡N} \xrightarrow[\text{H}_2\text{O}]{\text{Cu}_2\text{O, Cu}^{2+}} \text{Ar—OH}$$

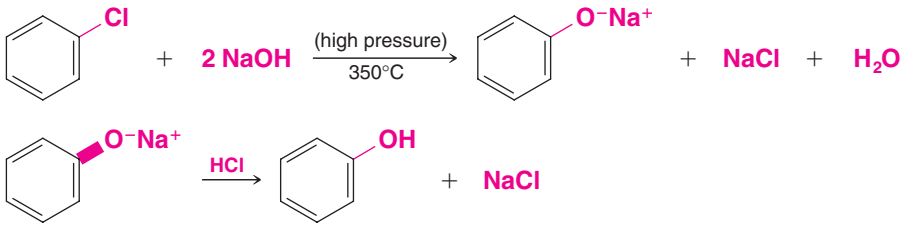
Specific Example

2-Bromo-4-methylphenol (80–92%)

21.4B Industrial Syntheses

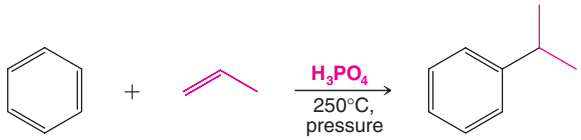
Phenol is a highly important industrial chemical; it serves as the raw material for a large number of commercial products ranging from aspirin to a variety of plastics. Worldwide production of phenol (which in industry is sometimes called carbolic acid) is more than 3 million tons per year. Several methods have been used to synthesize phenol commercially.

1. **Hydrolysis of Chlorobenzene (Dow Process).** In this process chlorobenzene is heated at 350°C (under high pressure) with aqueous sodium hydroxide. The reaction produces sodium phenoxide, which, on acidification, yields phenol. The mechanism for the reaction probably involves the formation of benzyne (Section 21.11B).



2. **From Cumene Hydroperoxide.** This process illustrates industrial chemistry at its best. Overall, it is a method for converting two relatively inexpensive organic compounds—benzene and propene—into two more valuable ones—phenol and acetone. The only other substance consumed in the process is oxygen from air. Most of the worldwide production of phenol is now based on this method. The synthesis begins with the Friedel–Crafts alkylation of benzene with propene to produce cumene (isopropylbenzene):

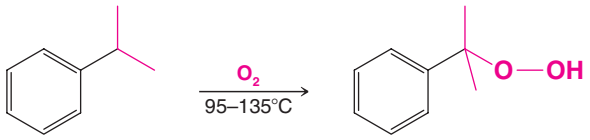
Reaction 1



Cumene

Then cumene is oxidized to cumene hydroperoxide:

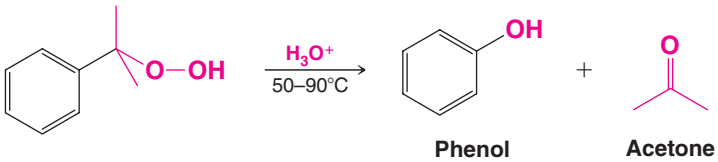
Reaction 2



Cumene hydroperoxide

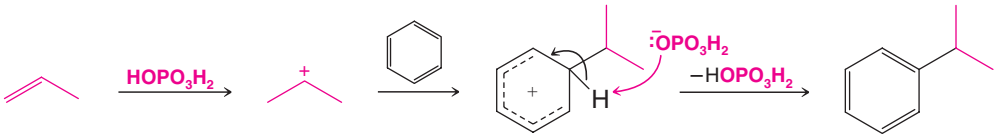
Finally, when treated with 10% sulfuric acid, cumene hydroperoxide undergoes a hydrolytic rearrangement that yields phenol and acetone:

Reaction 3



Phenol **Acetone**

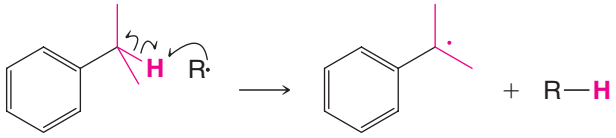
The mechanism of each of the reactions in the synthesis of phenol from benzene and propene via cumene hydroperoxide requires some comment. The first reaction is a familiar one. The isopropyl cation generated by the reaction of propene with the acid (H_3PO_4) alkylates benzene in a typical Friedel–Crafts electrophilic aromatic substitution:



The second reaction is a radical chain reaction. A radical initiator abstracts the benzylic hydrogen atom of cumene, producing a 3° benzylic radical. Then a chain reaction with oxygen, which exists as a paramagnetic diradical in the ground state, produces cumene hydroperoxide:

Chain Initiation

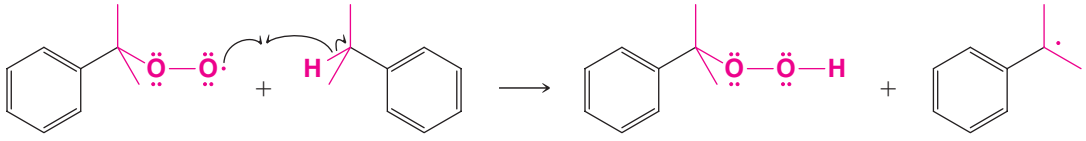
Step 1



Chain Propagation

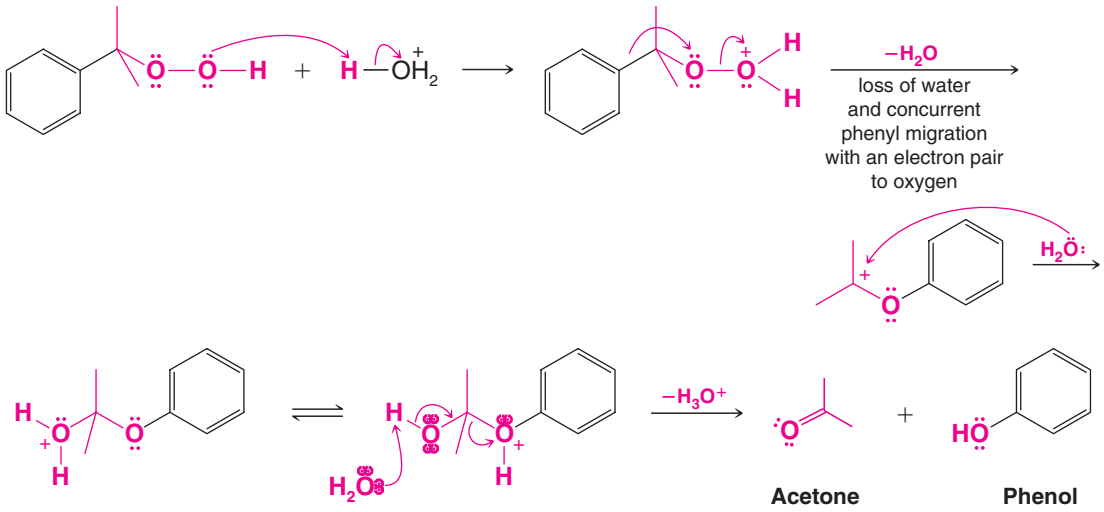
Step 2

Step 3



The reaction continues with steps 2, 3, 2, 3, etc.

The third reaction—the hydrolytic rearrangment—resembles the carbocation rearrangements that we have studied before. In this instance, however, the rearrangement involves the migration of a phenyl group to *a cationic oxygen atom*. Phenyl groups have a much greater tendency to migrate to a cationic center than do methyl groups. The following equations show all the steps of the mechanism.



The second and third steps of the mechanism may actually take place at the same time; that is, the loss of H_2O and the migration of C_6H_5— may be concerted.

21.5 Reactions of Phenols as Acids

21.5A Strength of Phenols as Acids

Although phenols are structurally similar to alcohols, they are much stronger acids. The pK_a values of most alcohols are of the order of 18. However, as we see in Table 21.1, the pK_a values of phenols are smaller than 11.

TABLE 21.1 Acidity Constants of Phenols

Name	pK_a (in H_2O at 25°C)	Name	pK_a (in H_2O at 25°C)
Phenol	9.89	3-Nitrophenol	8.28
2-Methylphenol	10.20	4-Nitrophenol	7.15
3-Methylphenol	10.01	2,4-Dinitrophenol	3.96
4-Methylphenol	10.17	2,4,6-Trinitrophenol	0.38
2-Chlorophenol	8.11	(picric acid)	
3-Chlorophenol	8.80	1-Naphthol	9.31
4-Chlorophenol	9.20	2-Naphthol	9.55
2-Nitrophenol	7.17		

Let us compare two *superficially* similar compounds, cyclohexanol and phenol:

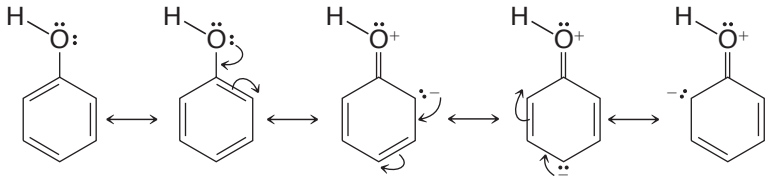
Cyclohexanol $pK_a = 18$ **Phenol** $pK_a = 9.89$

Although phenol is a weak acid when compared with a carboxylic acid such as acetic acid ($pK_a = 4.76$), phenol is a much stronger acid than cyclohexanol (by a factor of eight pK_a units).

Experimental and theoretical results have shown that the greater acidity of phenol owes itself primarily to an electrical charge distribution in phenol that causes the —OH oxygen to be more positive; therefore, the proton is held less strongly. In effect, the benzene ring of phenol acts as if it were an electron-withdrawing group when compared with the cyclohexane ring of cyclohexanol.

We can understand this effect by noting that the carbon atom which bears the hydroxyl group in phenol is sp^2 hybridized, whereas in cyclohexane it is sp^3 hybridized. Because of their greater *s* character, sp^2-hybridized carbon atoms are more electronegative than sp^3-hybridized carbon atoms (Section 3.8A).

Another factor influencing the electron distribution may be the contributions to the overall resonance hybrid of phenol made by structures **2–4**. Notice that the effect of these structures is to withdraw electrons from the hydroxyl group and to make the oxygen positive:

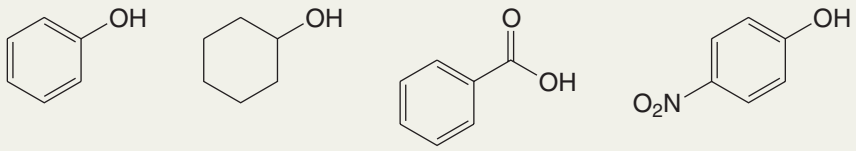


Resonance structures for phenol

An alternative explanation for the greater acidity of phenol relative to cyclohexanol can be based on similar resonance structures for the phenoxide ion. Unlike the structures for phenol, **2–4**, resonance structures for the phenoxide ion do not involve charge separation. According to resonance theory, such structures should stabilize the phenoxide ion more than structures **2–4** stabilize phenol. (No resonance structures can be written for cyclohexanol or its anion, of course.) Greater stabilization of the phenoxide ion (the conjugate base) than of phenol (the acid) has an acid-strengthening effect.

Solved Problem 21.1

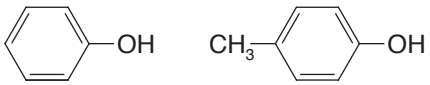
Rank the following compounds in order of increasing acidity.



STRATEGY AND ANSWER Alcohols are less acidic than phenols, and phenols are less acidic than carboxylic acids. An electron-withdrawing group increases the acidity of a phenol relative to phenol itself. Thus the order of increasing acidity among these examples is cyclohexanol <phenol <4-nitrophenol <benzoic acid.

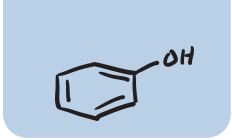
Review Problem 21.1

If we examine Table 21.1, we find that the methylphenols (cresols) are less acidic than phenol itself. For example,



Phenol $pK_a = 9.89$ **4-Methylphenol** $pK_a = 10.17$

This behavior is characteristic of phenols bearing electron-releasing groups. Provide an explanation.

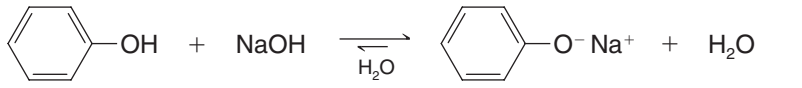


Review Problem 21.2

If we examine Table 21.1, we see that phenols having electron-withdrawing groups (Cl— or O_2N—) attached to the benzene ring are more acidic than phenol itself. Account for this trend on the basis of resonance and inductive effects. Your answer should also explain the large acid-strengthening effect of nitro groups, an effect that makes 2,4,6-trinitrophenol (also called *picric acid*) so exceptionally acidic ($pK_a = 0.38$) that it is more acidic than acetic acid ($pK_a = 4.76$).

21.5B Distinguishing and Separating Phenols from Alcohols and Carboxylic Acids

Because phenols are more acidic than water, the following reaction goes essentially to completion and produces water-soluble sodium phenoxide:



Stronger acid $pK_a \cong 10$ **(slightly soluble)** | **Stronger base** | **Weaker base (soluble)** | **Weaker acid** $pK_a \cong 16$

The corresponding reaction of 1-hexanol with aqueous sodium hydroxide does not occur to a significant extent because 1-hexanol is a weaker acid than water:

Weaker acid $pK_a \cong 18$ **(very slightly soluble)** | **Weaker base** | **Stronger base** | **Stronger acid** $pK_a \cong 16$

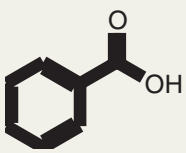
The fact that phenols dissolve in aqueous sodium hydroxide, whereas most alcohols with six carbon atoms or more do not, gives us a convenient means for distinguishing and separating phenols from most alcohols. (Alcohols with five carbon atoms or fewer are quite soluble in water—some are infinitely so—and so they dissolve in aqueous sodium hydroxide even though they are not converted to sodium alkoxides in appreciable amounts.)

Most phenols, however, are not soluble in aqueous sodium bicarbonate ($NaHCO_3$), but carboxylic acids are soluble. Thus, aqueous $NaHCO_3$ provides a method for distinguishing and separating most phenols from carboxylic acids.

Solved Problem 21.2

Assume that each of the following mixtures was added to a flask or a separatory funnel that contained diethyl ether (as an organic solvent) and mixed well. In which layer (diethyl ether or water) would the organic compound predominate in each case, and in what form would it exist (in its neutral form or as its conjugate base)?

(a)

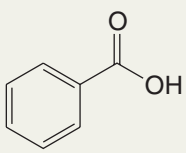


+ aqueous $NaHCO_3$

(b)

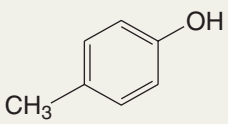
+ aqueous $NaHCO_3$

(c)



+ aqueous NaOH

(d)



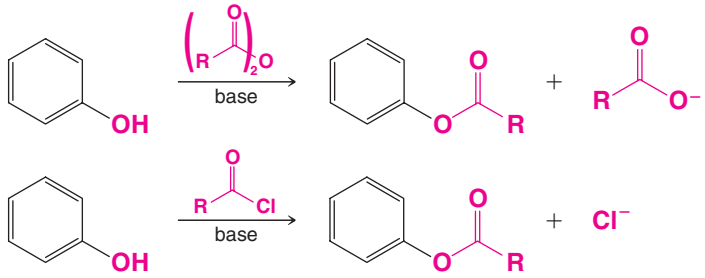
+ aqueous NaOH

STRATEGY AND ANSWER Sodium bicarbonate will remove a proton from a carboxylic acid to form a water-soluble carboxylate salt, but sodium bicarbonate will not remove a proton from a typical phenol. Sodium hydroxide will remove a proton from both a carboxylic acid and a phenol to form salts in each case. Thus, in (a) benzoic acid will be found in the water layer as its sodium salt, whereas in (b) 4-methylphenol will remain in its neutral form and be found predominantly in the ether layer. In (c) and (d) both benzoic acid and 4-methylphenol will be found in the aqueous layer as their corresponding salts.

Review Problem 21.3 Your laboratory instructor gives you a mixture of 4-methylphenol, benzoic acid, and toluene. Assume that you have available common laboratory acids, bases, and solvents and explain how you would proceed to separate this mixture by making use of the solubility differences of its components.

21.6 Other Reactions of the O—H Group of Phenols

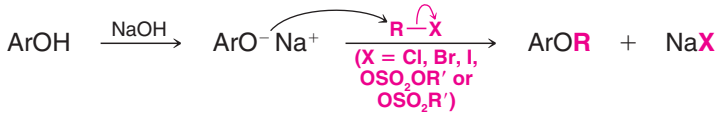
Phenols react with carboxylic acid anhydrides and acid chlorides to form esters. These reactions are quite similar to those of alcohols (Section 17.7).



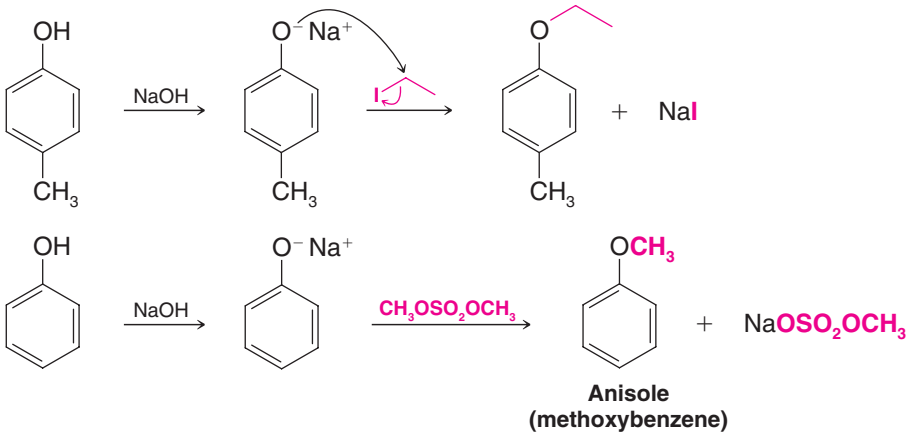
21.6A Phenols in the Williamson Synthesis

Phenols can be converted to ethers through the Williamson synthesis (Section 11.11B). Because phenols are more acidic than alcohols, they can be converted to sodium phenoxides through the use of sodium hydroxide (rather than sodium hydride or metallic sodium, the reagents used to convert alcohols to alkoxide ions).

General Reaction



Specific Examples



21.7 Cleavage of Alkyl Aryl Ethers

We learned in Section 11.12A that when dialkyl ethers are heated with excess concentrated HBr or HI, the ethers are cleaved and alkyl halides are produced from both alkyl groups:

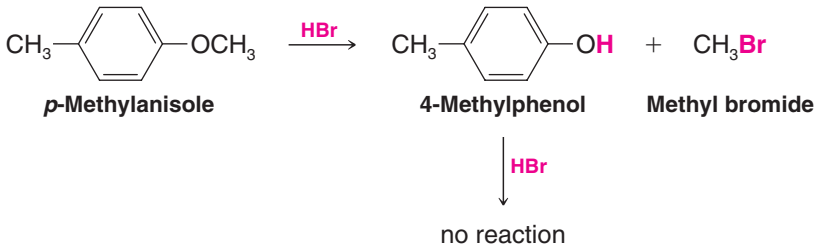
$$R{-}O{-}R' \xrightarrow[\text{heat}]{\text{concd HX}} R{-}X + R'{-}X + H_2O$$

When alkyl aryl ethers react with strong acids such as HI and HBr, the reaction produces an alkyl halide and a phenol. The phenol does not react further to produce an aryl halide because the phenol carbon–oxygen bond is very strong and because phenyl cations do not form readily.

General Reaction

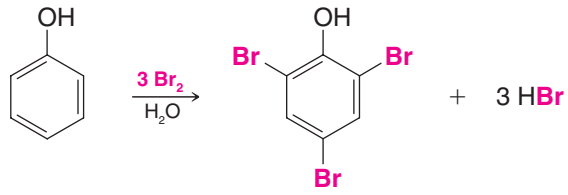
$$Ar{-}O{-}R \xrightarrow[\text{heat}]{\text{concd HX}} Ar{-}OH + R{-}X$$

Specific Example



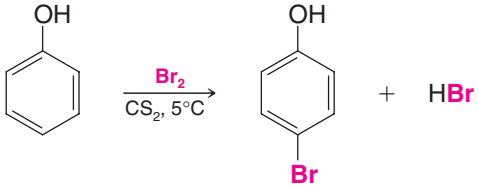
21.8 Reactions of the Benzene Ring of Phenols

Bromination The hydroxyl group is a powerful activating group—and an ortho–para director—in **electrophilic aromatic substitutions**. Phenol itself reacts with bromine in aqueous solution to yield 2,4,6-tribromophenol in nearly quantitative yield. Note that a Lewis acid is not required for the bromination of this highly activated ring:



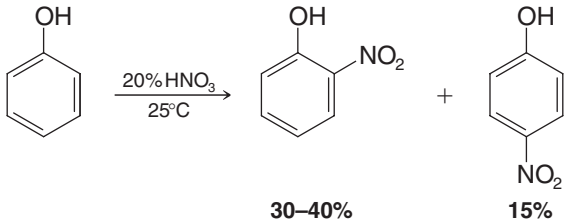
2,4,6-Tribromophenol (~100%)

Monobromination of phenol can be achieved by carrying out the reaction in carbon disulfide at a low temperature, conditions that reduce the electrophilic reactivity of bromine. The major product is the para isomer:

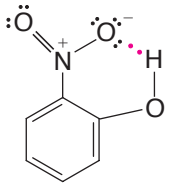


***p*-Bromophenol (80–84%)**

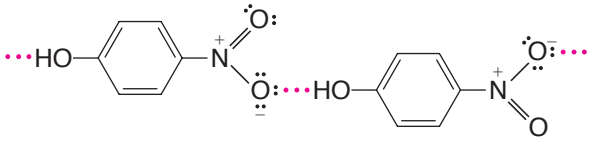
Nitration Phenol reacts with dilute nitric acid to yield a mixture of *o*- and *p*-nitrophenol:



Although the yield is relatively low (because of oxidation of the ring), the ortho and para isomers can be separated by steam distillation. *o*-Nitrophenol is the more volatile isomer because its hydrogen bonding (see the following structures) is *intramolecular*. *p*-Nitrophenol is less volatile because *intermolecular* hydrogen bonding causes association among its molecules. Thus, *o*-nitrophenol passes over with the steam, and *p*-nitrophenol remains in the distillation flask.

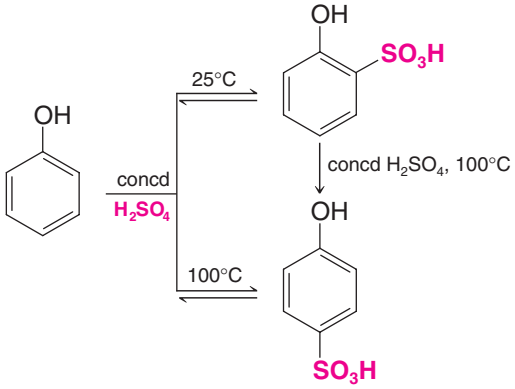


***o*-Nitrophenol**
(more volatile because of intramolecular hydrogen bonding)



***p*-Nitrophenol**
(less volatile because of intermolecular hydrogen bonding)

Sulfonation Phenol reacts with concentrated sulfuric acid to yield mainly the ortho-sulfonated product if the reaction is carried out at 25°C and mainly the para-sulfonated product at 100°C. This is another example of thermodynamic versus kinetic control of a reaction (Section 13.10A):



Major product under kinetic control

Major product under thermodynamic control

Solved Problem 21.3

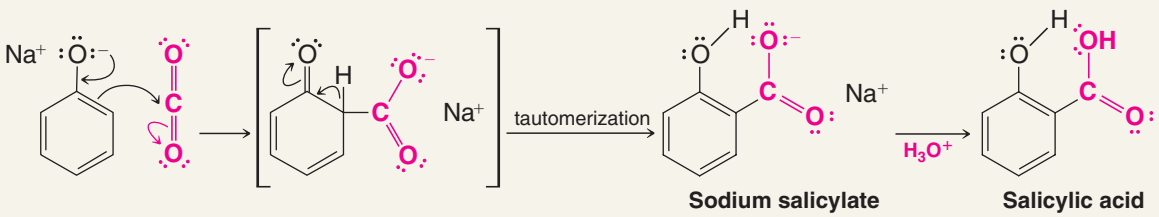
Consider the sulfonation reactions of phenol shown above. **(a)** Which sulfonation product is more stable, ortho or para? **(b)** For which sulfonation product is the free energy of activation lower?

ANSWER **(a)** The para-sulfonated phenol is more stable. We know this because at the higher temperature, where the reaction is under equilibrium control, it is the major product. **(b)** The free energy of activation is lower for ortho substitution. We know this because at the lower temperature, where the reaction is under kinetic control, it is formed faster.

Kolbe Reaction The phenoxide ion is even more susceptible to electrophilic aromatic substitution than phenol itself. (Why?) Use is made of the high reactivity of the phenoxide ring in a reaction called the *Kolbe reaction*. In the Kolbe reaction carbon dioxide acts as the electrophile.

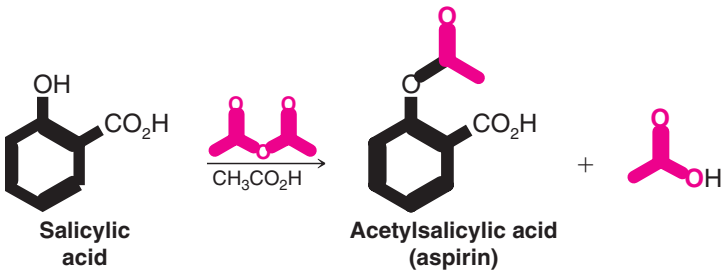
A MECHANISM FOR THE REACTION

The Kolbe Reaction



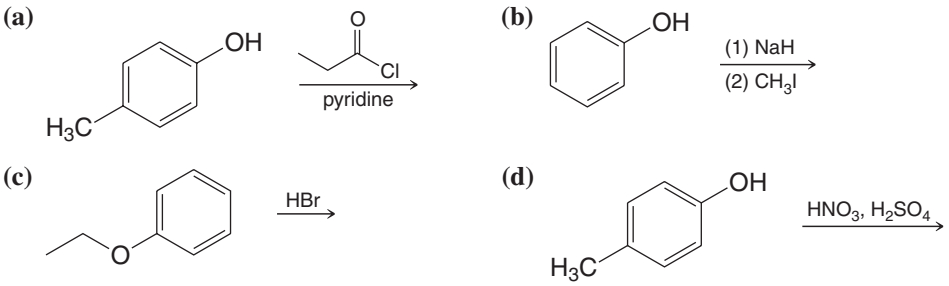
The Kolbe reaction is usually carried out by allowing sodium phenoxide to absorb carbon dioxide and then heating the product to 125°C under a pressure of several atmospheres of carbon dioxide. The unstable intermediate undergoes a proton shift (a keto–enol tautomerization; see Section 18.2) that leads to sodium salicylate. Subsequent acidification of the mixture produces *salicylic acid*.

Reaction of salicylic acid with acetic anhydride yields the widely used pain reliever *aspirin*:



Review Problem 21.4

Predict the products of each of the following reactions.

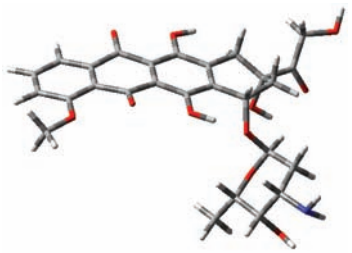


THE CHEMISTRY OF . . .

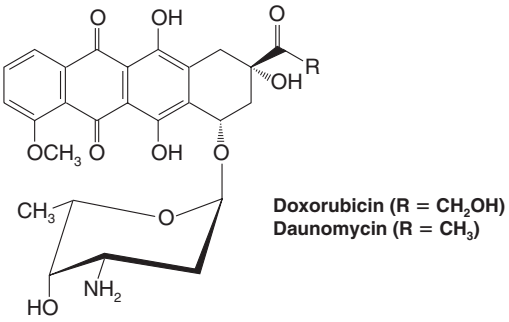
Polyketide Anticancer Antibiotic Biosynthesis

Doxorubicin (also known as adriamycin) is a highly potent anticancer drug that contains phenol functional groups. It is effective against many forms of cancer, including tumors of the ovaries, breast, bladder, and lung, as well as against Hodgkin's disease and other acute leukemias. Doxorubicin is a member of the anthracycline family of antibiotics.

Another member of the family is daunomycin. Both of these antibiotics are produced in strains of *Streptomyces* bacteria by a pathway called polyketide biosynthesis.



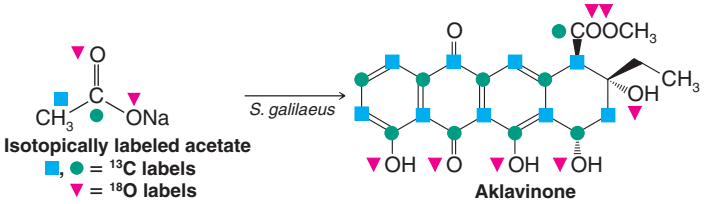
A molecular model of doxorubicin.



Doxorubicin (R = CH_2OH)
Daunomycin (R = CH_3)

Isotopic labeling experiments have shown that daunomycin is synthesized in *Streptomyces galilaeus* from a tetracyclic precursor called aklavinone. Aklavinone, in turn, is synthesized from acetate. When *S. galilaeus* is grown in a medium containing acetate labeled with carbon-13 and oxygen-18, the aklavinone produced has isotopic labels in the positions indicated below. Notice that oxygen atoms occur at alternate carbons in several places around the structure, consistent with the linking of acetate units in head-to-tail fashion. This is typical of aromatic polyketide biosynthesis.

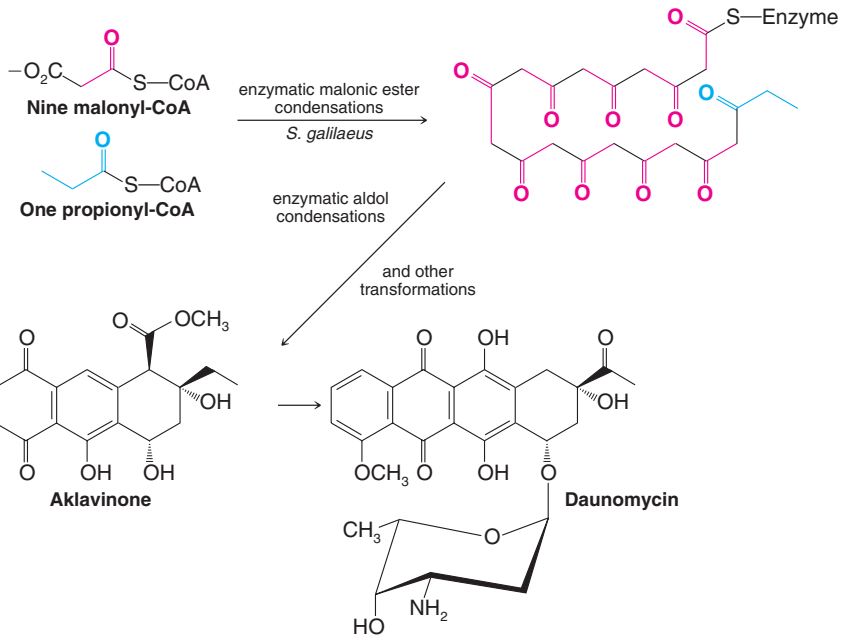
This and other information show that nine C_2 units from malonyl-coenzyme A and one C_3 unit from propionyl-coenzyme A condense to form the linear polyketide intermediate shown below. These units are joined by acylation reactions that are the biosynthetic equivalent of the *malonic ester synthesis* we studied in Section 18.7. These reactions



Isotopically labeled acetate
■, ● = ^{13}C labels
▼ = ^{18}O labels

Aklavinone

are also similar to the acylation steps we saw in fatty acid biosynthesis (Special Topic E). Once formed, the linear polyketide cyclizes by enzymatic reactions akin to intramolecular *aldol additions and dehydrations* (Section 19.6). These steps form the tetracyclic core of aklavinone. Phenolic hydroxyl groups in aklavinone arise by enolization of ketone carbonyl groups present after the aldol condensation steps. Several other transformations ultimately lead to daunomycin:



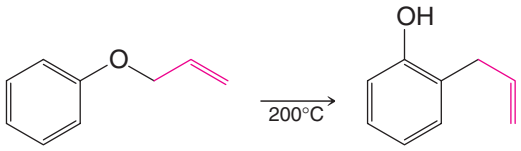
Aklavinone

Daunomycin

There are many examples of important biologically active molecules formed by polyketide biosynthesis. Aureomycin and Terramycin (Section 21.2) are examples of other aromatic polyketide antibiotics. Erythromycin (Section 17.7C) and aflatoxin, a carcinogen (Section 11.14), are polyketides from other pathways.

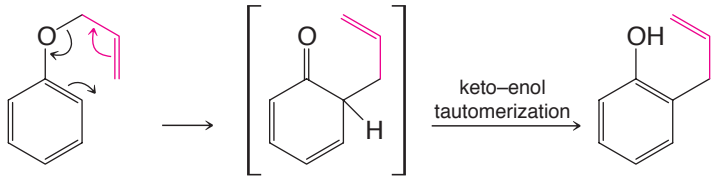
21.9 The Claisen Rearrangement

Heating allyl phenyl ether to 200°C effects an intramolecular reaction called a **Claisen rearrangement**. The product of the rearrangement is *o*-allylphenol:



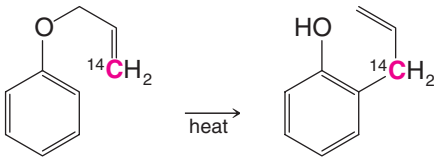
Allyl phenyl ether **o-Allylphenol**

The reaction takes place through a **concerted rearrangement** in which the bond between C3 of the allyl group and the ortho position of the benzene ring forms at the same time that the carbon–oxygen bond of the allyl phenyl ether breaks. The product of this rearrangement is an unstable intermediate that, like the unstable intermediate in the Kolbe reaction (Section 21.8), undergoes a proton shift (a keto–enol tautomerization, see Section 18.2) that leads to the *o*-allylphenol:



Unstable intermediate

That only C3 of the allyl group becomes bonded to the benzene ring was demonstrated by carrying out the rearrangement with allyl phenyl ether containing ^{14}C at C3. All of the product of this reaction had the labeled carbon atom bonded to the ring:



Only product

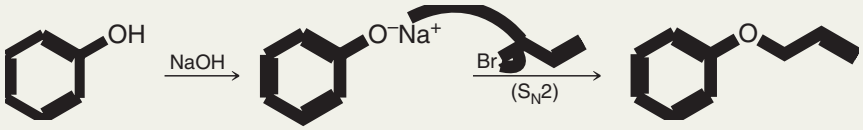
Review Problem 21.5

The labeling experiment just described eliminates from consideration a mechanism in which the allyl phenyl ether dissociates to produce an allyl cation (Section 13.4) and a phenoxide ion, which then subsequently undergo a Friedel–Crafts alkylation (Section 15.6) to produce the *o*-allylphenol. Explain how this alternative mechanism can be discounted by showing the product (or products) that would result from it.

Solved Problem 21.4

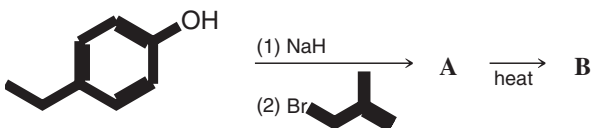
Show how you could synthesize allyl phenyl ether from phenol and allyl bromide.

STRATEGY AND ANSWER Use a Williamson ether synthesis (Section 21.6A).

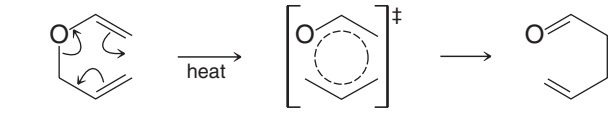


Review Problem 21.6

What are compounds **A** and **B** in the following sequence?

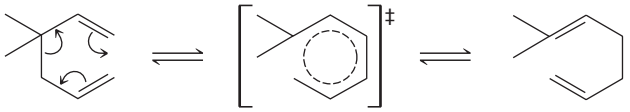


A Claisen rearrangement also takes place when allyl vinyl ethers are heated. For example,



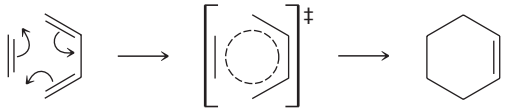
Allyl vinyl ether **Aromatic transition state** **4-Pentenal**

The transition state for the Claisen rearrangement involves a cycle of six electrons. Having six electrons suggests that the transition state has aromatic character (Section 14.7). Other reactions of this general type are known, and they are called **pericyclic reactions**. Another similar pericyclic reaction is the **Cope rearrangement** shown here:



3,3-Dimethyl-1,5-hexadiene **Aromatic transition state** **2-Methyl-2,6-heptadiene**

The Diels–Alder reaction (Section 13.11) is also a pericyclic reaction. The transition state for the Diels–Alder reaction also involves six electrons:

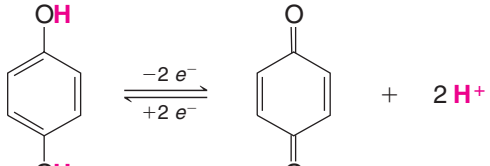


Aromatic transition state

The mechanism of the Diels–Alder reaction is discussed further in Special Topic H.

21.10 *Quinones*

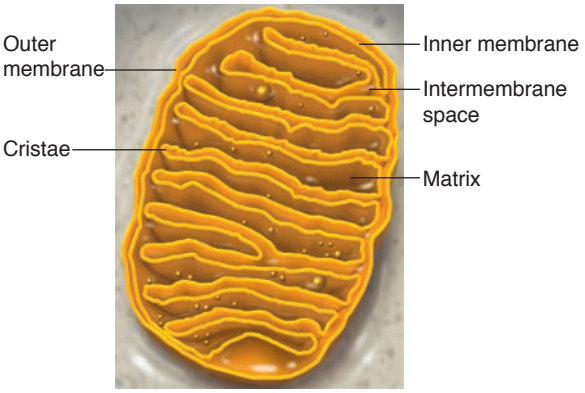
Oxidation of hydroquinone (1,4-benzenediol) produces a compound known as *p*-benzoquinone. The oxidation can be brought about by mild oxidizing agents, and, overall, the oxidation amounts to the removal of a pair of electrons ($2\ e^-$) and two protons from hydroquinone. (Another way of visualizing the oxidation is as the loss of a hydrogen molecule, H:H, making it a dehydrogenation.)



Hydroquinone ***p*-Benzoquinone**

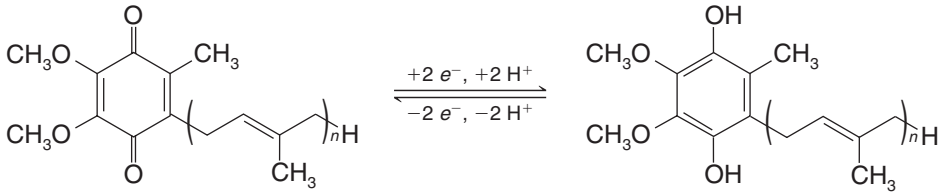
This reaction is reversible; *p*-benzoquinone is easily reduced by mild reducing agents to hydroquinone.

Nature makes much use of this type of reversible oxidation–reduction to transport a pair of electrons from one substance to another in enzyme-catalyzed reactions. Important compounds in this respect are the compounds called **ubiquinones** (from *ubiquitous* + quinone—these quinones are found within the inner mitochondrial membrane of every living cell). Ubiquinones are also called coenzymes Q (CoQ).



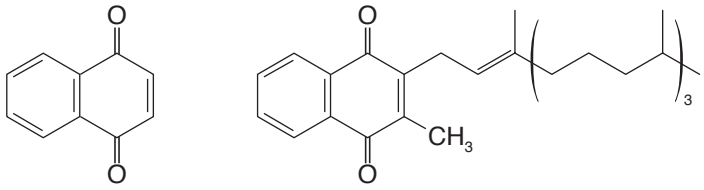
Cross section of a mitochondrion.

Ubiquinones have a long, isoprene-derived side chain (see Special Topic E and Section 23.3). Ten isoprene units are present in the side chain of human ubiquinones. This part of their structure is highly nonpolar, and it serves to solubilize the ubiquinones within the hydrophobic bilayer of the mitochondrial inner membrane. Solubility in the membrane environment facilitates their lateral diffusion from one component of the electron transport chain to another. In the electron transport chain, ubiquinones function by accepting two electrons and two hydrogen atoms to become a hydroquinone. The hydroquinone form carries the two electrons to the next acceptor in the chain:

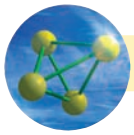


Ubiquinones (n = 6–10) (coenzymes Q) **Ubiquinol (hydroquinone form)**

Vitamin K_1, the important dietary factor that is instrumental in maintaining the coagulant properties of blood, contains a 1,4-naphthoquinone structure:



1,4-Naphthoquinone **Vitamin K_1**



THE CHEMISTRY OF . . .

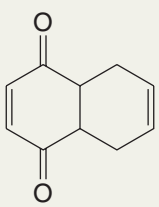
The Bombardier Beetle's Noxious Spray

The bombardier beetle defends itself by spraying a jet stream of hot (100°C), noxious *p*-benzoquinones at an attacker. The beetle mixes *p*-hydroquinones and hydrogen peroxide from one abdominal reservoir with enzymes from another reservoir. The enzymes convert hydrogen peroxide to oxygen, which in turn oxidizes the *p*-hydroquinones to *p*-benzoquinones and explosively propels the irritating spray at the attacker. Photos by T. Eisner and D. Aneshansley (Cornell University) have shown that the amazing bombardier beetle can direct its spray in virtually any direction, even parallel over its back, to ward off a predator.

Bombardier beetle in the process of spraying.

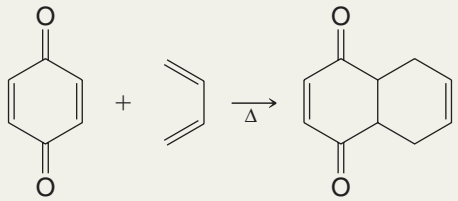
Solved Problem 21.5

Outline a synthesis of the following compound.



(continued on the next page)

STRATEGY AND ANSWER The presence of a cyclohexane ring with a double bond in it suggests that the compound could be made by a Diels–Alder reaction. Suitable reactants here would be *p*-benzoquinone as the dienophile and 1,3-butadiene as the diene.



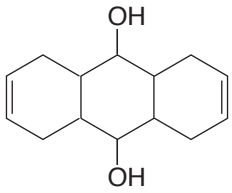
Review Problem 21.7

p-Benzoquinone and 1,4-naphthoquinone act as dienophiles in Diels–Alder reactions. Give the structures of the products of the following reactions:

(a) 1,4-Naphthoquinone + butadiene **(b)** *p*-Benzoquinone + cyclopentadiene

Review Problem 21.8

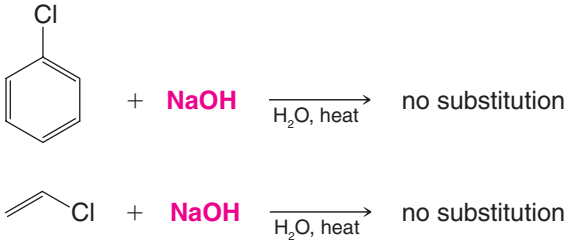
Outline a possible synthesis of the following compound.



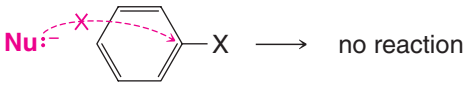
21.11 *Aryl Halides and Nucleophilic Aromatic Substitution*

- Simple aryl halides, like vinylic halides (Section 6.14A), are relatively unreactive toward nucleophilic substitution under conditions that would give rapid nucleophilic substitution with alkyl halides.

Chlorobenzene, for example, can be boiled with sodium hydroxide for days without producing a detectable amount of phenol (or sodium phenoxide).* Similarly, when vinyl chloride is heated with sodium hydroxide, no substitution occurs:



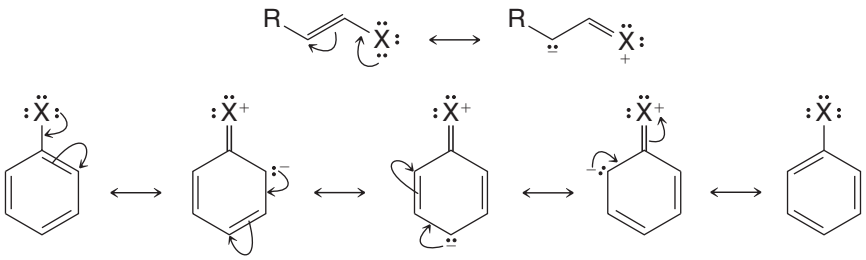
We can understand this lack of reactivity on the basis of several factors. The benzene ring of an aryl halide prevents back-side attack in an S_N2 reaction:



*The Dow process for making phenol by substitution (Section 21.4B) requires extremely high temperature and pressure to effect the reaction. These conditions are not practical in the laboratory.

Phenyl cations are very unstable; thus S_N1 reactions do not occur. The carbon–halogen bonds of aryl (and vinylic) halides are shorter and stronger than those of alkyl, allylic, and benzylic halides. Stronger carbon–halogen bonds mean that bond breaking by either an S_N1 or S_N2 mechanism will require more energy.

Two effects make the carbon–halogen bonds of aryl and vinylic halides shorter and stronger: (1) The carbon of either type of halide is sp^2 hybridized, and therefore the electrons of the carbon orbital are closer to the nucleus than those of an sp^3-hybridized carbon. (2) Resonance of the type shown here strengthens the carbon–halogen bond by giving it *double-bond character*:

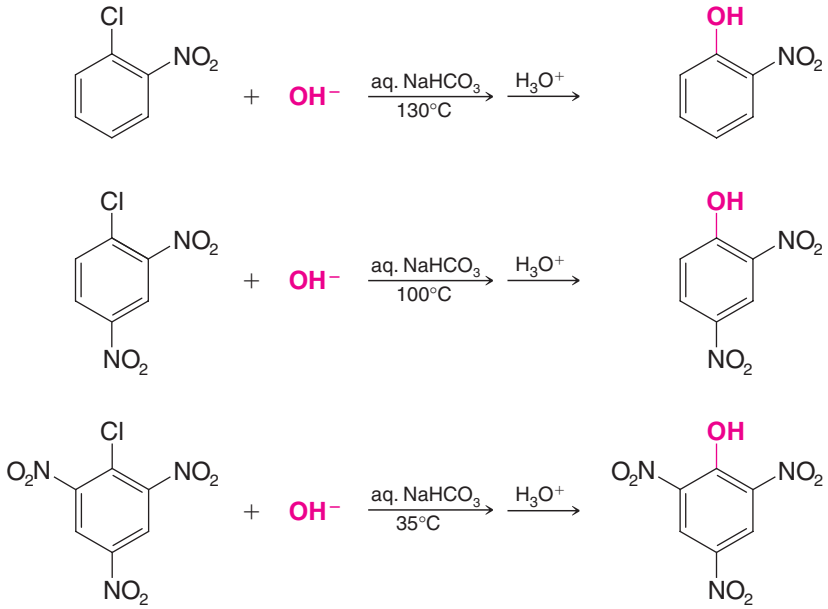


Having said all this, we shall find in the next two subsections that *aryl halides can be remarkably reactive toward nucleophiles* if they bear certain substituents or when we allow them to react under the proper conditions.

21.11A Nucleophilic Aromatic Substitution by Addition–Elimination: The S_NAr Mechanism

Nucleophilic substitution reactions of aryl halides *do* occur readily when an electronic factor makes the aryl carbon bonded to the halogen susceptible to nucleophilic attack.

- Nucleophilic aromatic substitution can occur when strong electron-withdrawing groups are ortho or para to the halogen atom:



We also see in these examples that the temperature required to bring about the reaction is related to the number of ortho or para nitro groups. Of the three compounds, *o*-nitrochlorobenzene requires the highest temperature (*p*-nitrochlorobenzene reacts at 130°C as well) and 2,4,6-trinitrochlorobenzene requires the lowest temperature.

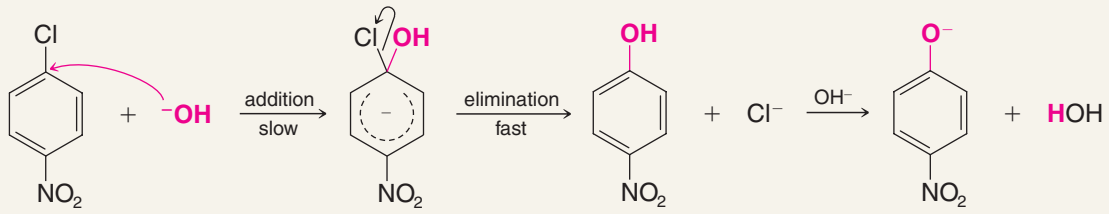
A meta-nitro group does not produce a similar activating effect. For example, *m*-nitrochlorobenzene gives no corresponding reaction.

- The mechanism that operates in these reactions is an *addition–elimination* mechanism involving the formation of a *carbanion* with delocalized electrons, called a **Meisenheimer intermediate**. The process is called **nucleophilic aromatic substitution (S_NAr)**.

In the first step of the following example, addition of a hydroxide ion to *p*-nitrochlorobenzene produces the carbanion; then elimination of a chloride ion yields the substitution product as the aromaticity of the ring is recovered.

A MECHANISM FOR THE REACTION

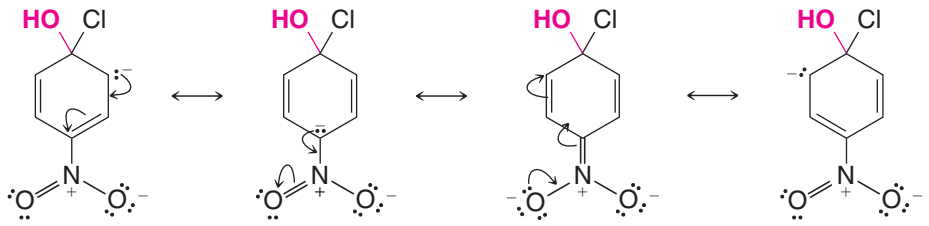
The S_NAr Mechanism



Carbanion (Meisenheimer intermediate)

Structures of the contributing resonance forms are shown below.

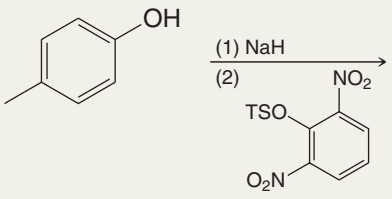
The carbanion is stabilized by *electron-withdrawing groups* in the positions ortho and para to the halogen atom. If we examine the following resonance structures for a Meisenheimer intermediate, we can see how:



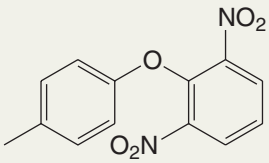
Especially stable (Negative charges are both on oxygen atoms.)

Solved Problem 21.6

What is the product of the following reaction?



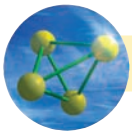
STRATEGY AND ANSWER NaH is a strong base that will convert 4-methylphenol to its phenoxide salt. 1-(*p*-Toluenesulfonyl)-2,6-dinitrobenzene contains both a good leaving group and two strong electron-withdrawing groups. Thus the likely reaction is a nucleophilic aromatic substitution (S_NAr), leading to the following diaryl ether.



Review Problem 21.9

1-Fluoro-2,4-dinitrobenzene is highly reactive toward nucleophilic substitution through an S_NAr mechanism. (In Section 24.5B we shall see how this reagent is used in the Sanger method for determining the structures of proteins.) What product would be formed when 1-fluoro-2,4-dinitrobenzene reacts with each of the following reagents?

(a) EtONa **(b)** NH_3 **(c)** $C_6H_5NH_2$ **(d)** EtSNa

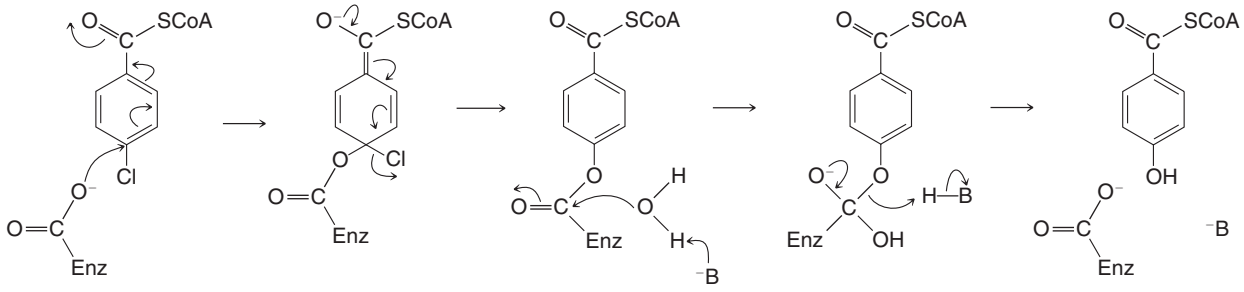


THE CHEMISTRY OF . . .

Bacterial Dehalogenation of a PCB Derivative

Polychlorinated biphenyls (PCBs) are compounds that were once used in a variety of electrical devices, industrial applications, and polymers. Their use and production were banned in 1979, however, owing to the toxicity of PCBs and their tendency to accumulate in the food chain.

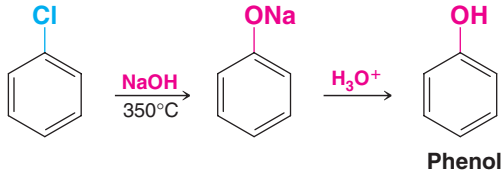
4-Chlorobenzoic acid is a degradation product of some PCBs. It is now known that certain bacteria are able to dehalogenate 4-chlorobenzoic acid by an enzymatic nucleophilic aromatic substitution reaction. The product is 4-hydroxybenzoic acid, and a mechanism for this enzyme-catalyzed process is shown here. The sequence begins with the thioester of 4-chlorobenzoic acid derived from coenzyme A (CoA):



Some key features of this enzymatic S_NAr mechanism are the following. The nucleophile that attacks the chlorinated benzene ring is a carboxylate anion of the enzyme. When the carboxylate attacks, positively charged groups within the enzyme stabilize the additional electron density that develops in the thioester carbonyl group of the Meisenheimer intermediate. Collapse of the Meisenheimer intermediate, with rearomatization of the ring and loss of the chloride ion, results in an intermediate where the substrate is covalently bonded to the enzyme as an ester. Hydrolysis of this ester linkage involves a water molecule whose nucleophilicity has been enhanced by a basic site within the enzyme. Hydrolysis of the ester releases 4-hydroxybenzoic acid and leaves the enzyme ready to catalyze another reaction cycle.

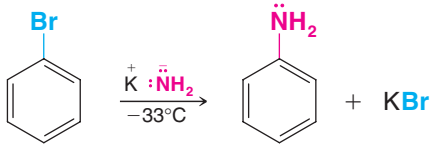
21.11B Nucleophilic Aromatic Substitution through an Elimination–Addition Mechanism: Benzyne

Although aryl halides such as chlorobenzene and bromobenzene do not react with most nucleophiles under ordinary circumstances, they do react under highly forcing conditions. Chlorobenzene can be converted to phenol by heating it with aqueous sodium hydroxide in a pressurized reactor at 350°C (Section 21.4B):



Phenol

Bromobenzene reacts with the very powerful base, $^-NH_2$, in liquid ammonia:



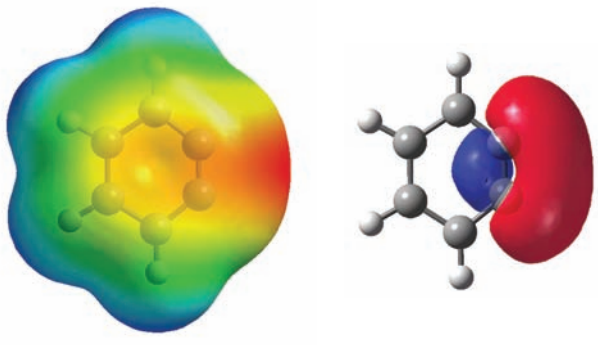
Aniline

- These reactions take place through an **elimination–addition mechanism** that involves the formation of a highly unstable intermediate called *benzyne* (or *dehydrobenzene*).

We can illustrate this mechanism with the reaction of bromobenzene and amide ion. In the first step (see the following mechanism), the amide ion initiates an elimination by abstracting one of the ortho protons because they are the most acidic. The negative charge that develops on the ortho carbon is stabilized by the inductive effect of the bromine. The anion then loses a bromide ion. This elimination produces the highly unstable, and thus highly reactive, **benzyne**. Benzyne then reacts with any available nucleophile (in this case, an amide ion) by a two-step addition reaction to produce aniline.

We can better understand the reactive and unstable nature of benzyne if we consider aspects of its electronic structure.

The calculated electrostatic potential map for benzyne, shown in Fig. 21.1*a*, shows the relatively greater negative charge at the edge of the ring, corresponding to the electron density from the additional π bond in benzyne. Figure 21.1*b* shows a schematic representation of the orbital associated with the additional π bond. We can see from these models that the orbitals of the additional π bond in benzyne lie in the same plane as the ring, perpendicular to the axis of the aromatic π system. We can also see in Fig. 21.1 that, because the carbon ring is not a perfect hexagon as in benzene, there is angle strain in the structure of benzyne. The distance between the carbons of the additional π bond in benzyne is shorter than between the other carbons, and the bond angles of the ring are therefore distorted from their ideal values. The result is that benzyne is highly unstable and highly reactive. Consequently, benzyne has never been isolated as a pure substance, but it has been detected and trapped in various ways.



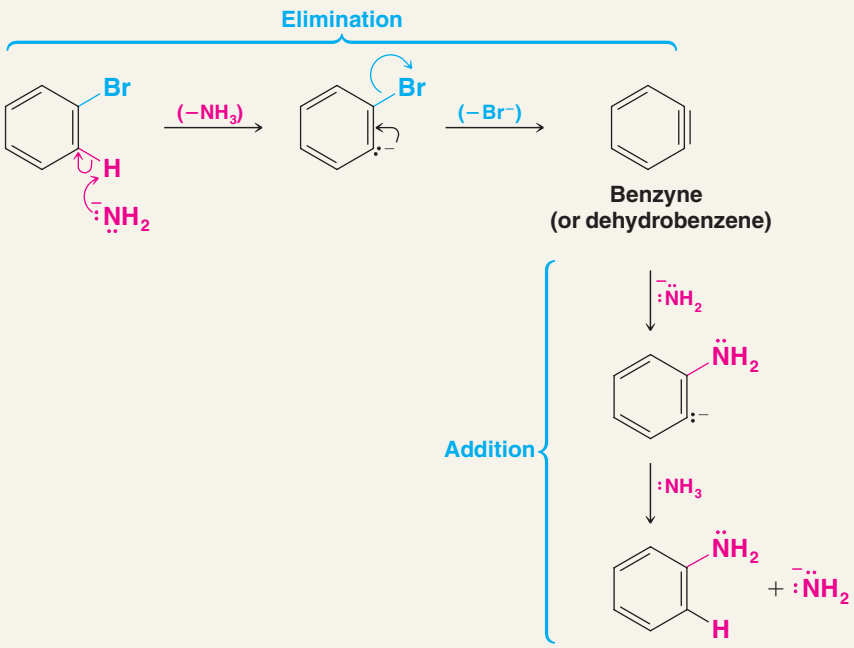
(*a*) (*b*)

Figure 21.1 *(a)* A calculated electrostatic potential map for benzyne shows the relatively greater negative charge (in red) at the edge of the ring, corresponding to electron density from the additional π bond in benzyne. *(b)* A schematic representation of the molecular orbital associated with the additional π bond in benzyne. (Red and blue indicate orbital phase, not charge distribution.) Note that the orientation of this orbital is in the same plane as the ring and perpendicular to the axis of the aromatic π system.

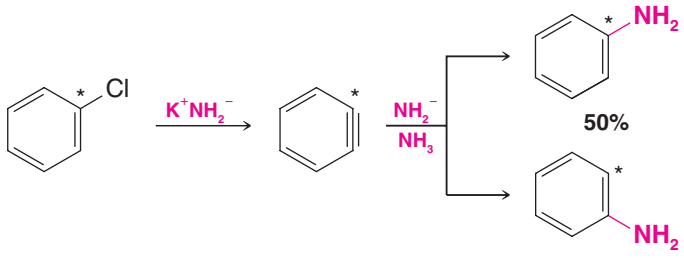
What, then, is some of the evidence for an elimination–addition mechanism involving benzyne in some nucleophilic aromatic substitutions?

A MECHANISM FOR THE REACTION

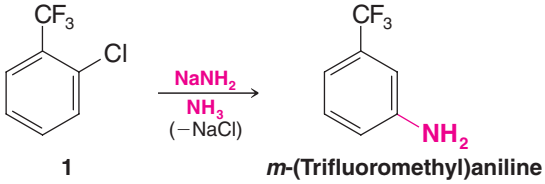
The Benzyne Elimination–Addition Mechanism



The first piece of clear-cut evidence was an experiment done by J. D. Roberts (Section 9.10) in 1953—one that marked the beginning of benzyne chemistry. Roberts showed that when ^{14}C-labeled (C*) chlorobenzene is treated with amide ion in liquid ammonia, the aniline that is produced has the label equally divided between the 1 and 2 positions. This result is consistent with the following elimination–addition mechanism but is, of course, not at all consistent with a direct displacement or with an addition–elimination mechanism. (Why?)

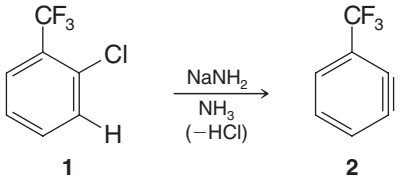


An even more striking illustration can be seen in the following reaction. When the ortho derivative **1** is treated with sodium amide, the only organic product obtained is *m*-(trifluoromethyl)aniline:



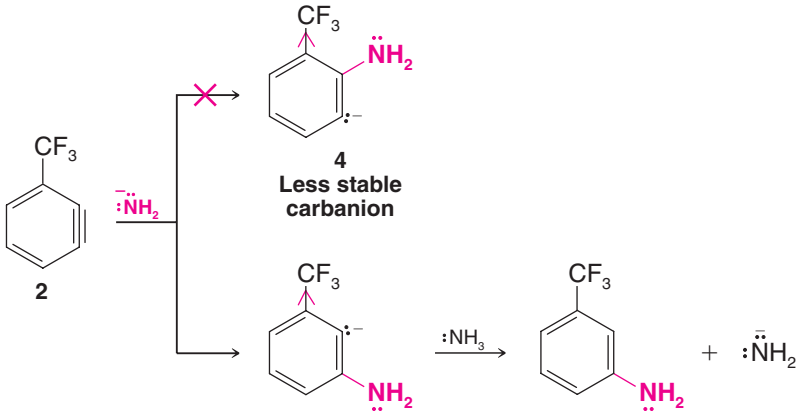
1 ***m*-(Trifluoromethyl)aniline**

This result can also be explained by an elimination–addition mechanism. The first step produces the benzyne **2**:



1 **2**

This benzyne then adds an amide ion in the way that produces the more stable carbanion **3** rather than the less stable carbanion **4**:

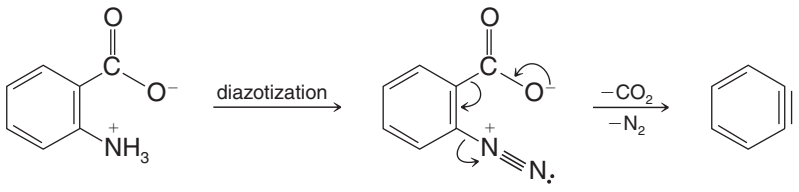


More stable carbanion (The negative charge is closer to the electronegative trifluoromethyl group.)

Carbanion **3** then accepts a proton from ammonia to form *m*-(trifluoromethyl)aniline.

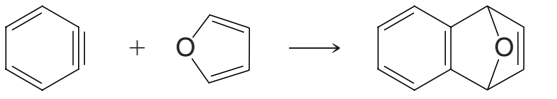
Carbanion **3** is more stable than **4** because the carbon atom bearing the negative charge is closer to the highly electronegative trifluoromethyl group. The trifluoromethyl group stabilizes the negative charge through its inductive effect. (Resonance effects are not important here because the sp^2 orbital that contains the electron pair does not overlap with the π orbitals of the aromatic system.)

Benzyne intermediates have been "trapped" through the use of Diels–Alder reactions. One convenient method for generating benzyne is the diazotization of anthranilic acid (2-aminobenzoic acid) followed by elimination of CO_2 and N_2:



Anthranilic acid **Benzyne (trapped *in situ*)**

When benzyne is generated in the presence of the diene *furan*, the product is a Diels–Alder adduct:



Benzyne (generated by an elimination reaction) **Furan** **Diels–Alder adduct**

In a fascinating application of host–guest chemistry (an area founded by the late D. Cram, and for which he shared the Nobel Prize in Chemistry in 1987), benzyne itself has been trapped at very low temperature inside a molecular container called a hemicarcerand. Under these conditions, R. Warmuth and Cram found that the incarcerated benzyne was sufficiently stabilized for its ^{1}H and ^{13}C NMR spectra to be recorded (see Fig. 21.2), before it ultimately underwent a Diels–Alder reaction with the container molecule.

Donald Cram shared the 1987 Nobel prize for his work on host–guest chemistry.

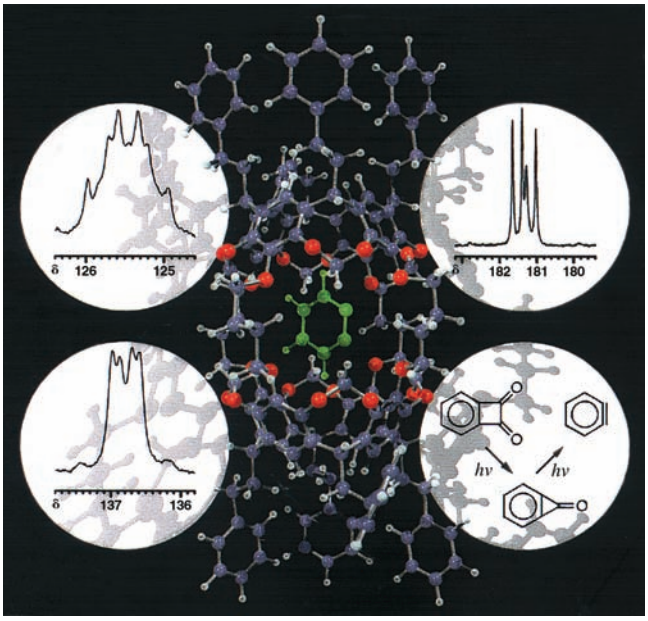
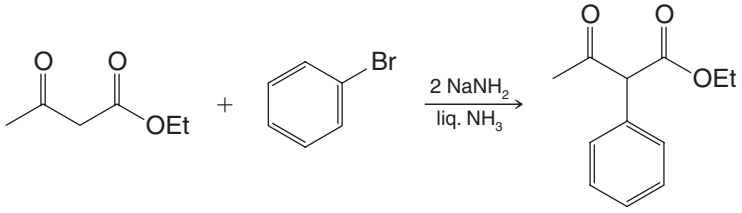


Figure 21.2 A molecular graphic of benzyne (green) trapped in a hemicarcerand. Images of ^{13}C NMR data from benzyne and a reaction used to synthesize it are shown in the white circles.

21.11C Phenylation

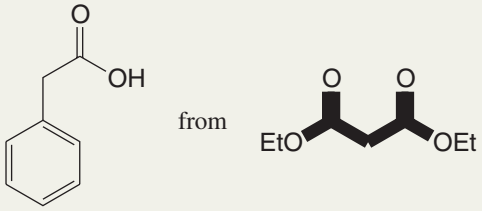
Reactions involving benzyne can be useful for formation of a carbon–carbon bond to a phenyl group (a process called phenylation). For example, if acetoacetic ester is treated with bromobenzene and two molar equivalents of sodium amide, phenylation of ethyl acetoacetate occurs. The overall reaction is as follows:



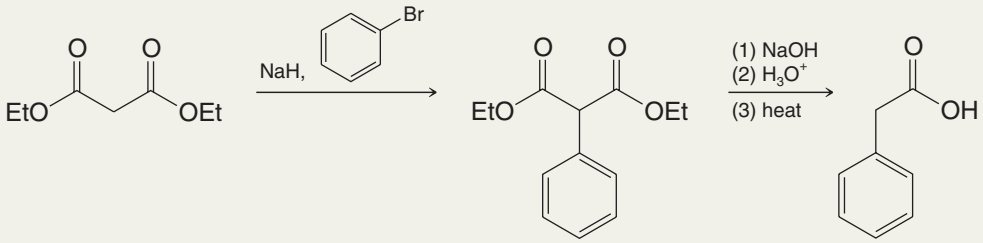
Malonic esters can be phenylated in an analogous way. This process is a useful complement to the alkylation reactions of acetoacetic and malonic esters that we studied in Chapter 18 because, as you may recall, substrates like bromobenzene are not susceptible to S_N2 reactions [see Section 6.14A and Review Problem 18.8(c)].

Solved Problem 21.7

Outline a synthesis of phenylacetic acid from diethyl malonate.



STRATEGY AND ANSWER Diethyl malonate must first be substituted at the α carbon by a phenyl group, and then hydrolyzed and decarboxylated. Introduction of the phenyl group requires involvement of a benzyne intermediate.



Review Problem 21.10 When *o*-chlorotoluene is subjected to the conditions used in the Dow process (i.e., aqueous NaOH at 350°C at high pressure), the products of the reaction are *o*-cresol and *m*-cresol. What does this result suggest about the mechanism of the Dow process?

Review Problem 21.11 When 2-bromo-1,3-dimethylbenzene is treated with sodium amide in liquid ammonia, no substitution takes place. This result can be interpreted as providing evidence for the elimination–addition mechanism. Explain how this interpretation can be given.

Review Problem 21.12 **(a)** Outline a step-by-step mechanism for the phenylation of acetoacetic ester by bromobenzene and two molar equivalents of sodium amide. (Why are two molar equivalents of $NaNH_2$ necessary?) **(b)** What product would be obtained by hydrolysis and decarboxylation of the phenylated acetoacetic ester? **(c)** How would you prepare 2-phenylpropanoic acid from malonic ester?

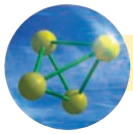
21.12 *Spectroscopic Analysis of Phenols and Aryl Halides*

Infrared Spectra Phenols show a characteristic absorption band (usually broad) arising from O—H stretching in the 3400–3600-cm^{-1} region. Phenols and aryl halides also show the characteristic absorptions that arise from their benzene rings (see Section 14.11C).

^{1}H NMR Spectra The hydroxylic proton of a phenol is more deshielded than that of an alcohol due to proximity of the benzene π electron ring current. The exact position of the O—H signal depends on the extent of hydrogen bonding and on whether the hydrogen bonding is *intermolecular* or *intramolecular*. The extent of intermolecular hydrogen bonding depends on the concentration of the phenol, and this strongly affects the position of the O—H signal. In phenol, itself, for example, the position of the O—H signal varies from δ 2.55 for pure phenol to δ 5.63 at 1% concentration in CCl_4. Phenols with strong intramolecular hydrogen bonding, such as salicylaldehyde, show O—H signals between δ 0.5 and δ 1.0, and the position of the signal varies only slightly with concentration. As with other protons that undergo exchange (Section 9.10), the identity of the O—H proton of a phenol can be determined by adding D_2O to the sample. The O—H proton undergoes rapid exchange with deuterium and the proton signal disappears. The aromatic protons of phenols and aryl halides give signals in the δ 7–9 region.

^{13}C NMR Spectra The carbon atoms of the aromatic ring of phenols and aryl halides appear in the region δ 135–170.

Mass Spectra Mass spectra of phenols often display a prominent molecular ion peak, $M^{+\cdot}$. Phenols that have a benzylic hydrogen produce an $M^{+\cdot} - 1$ peak that can be larger than the $M^{+\cdot}$ peak.



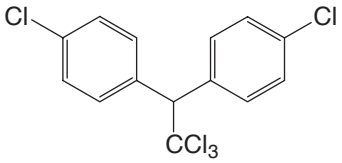
THE CHEMISTRY OF . . .

Aryl Halides: Their Uses and Environmental Concerns

Aryl Halides as Insecticides

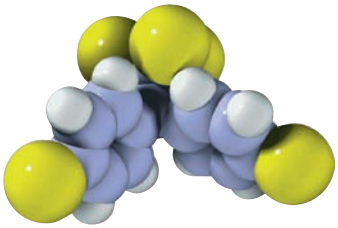
Insects, especially mosquitoes, fleas, and lice, have been responsible for innumerable human deaths throughout history. The bubonic plague or "black death" of medieval times that killed nearly one-third of Europe's population was borne by fleas. Malaria and yellow fever, diseases that were responsible for the loss of millions of lives in the twentieth century alone, are mosquito-borne diseases.

One compound widely known for its insecticidal properties and environmental effects is DDT [1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane].



DDT
[1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane]

From the early 1940s through the early 1970s, when its use was banned in the United States, vast quantities of DDT were sprayed over many parts of the world in an effort to destroy insects. These efforts rid large areas of the world of disease-carrying insects, especially those responsible for malaria, yellow fever, sleeping sickness (caused by tsetse flies), and typhus. Though it has since resurged, by 1970, malaria had been



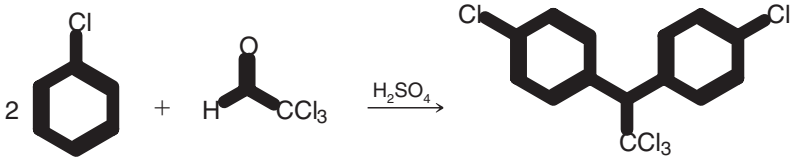
DDT.

largely eliminated from the developed world. According to estimates by the National Academy of Sciences, the use of DDT during that time had prevented more that 500 million deaths from malaria alone.

Eventually it began to become clear that the prodigious use of DDT had harmful side effects. Aryl halides are usually highly stable compounds that are only slowly destroyed by natural processes. As a result they remain in the environment for years; they are what we now call "persistent insecticides" or "hard insecticides." The U.S. Environmental Protection Agency banned the use of DDT beginning in 1973.

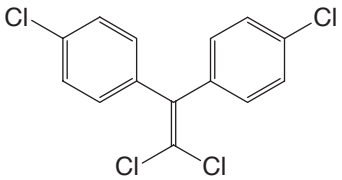
Aryl halides are also fat soluble and tend to accumulate in the fatty tissues of most animals. The food chain that runs from plankton to small fish to birds and to larger animals, including humans, tends to magnify the concentrations of aryl halides at each step.

The chlorohydrocarbon DDT is prepared from inexpensive starting materials, chlorobenzene and trichloroacetaldehyde. The reaction, shown here, is catalyzed by acid.



DDT
[1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane]

In nature the principal decomposition product of DDT is DDE.



DDE
[1,1-dichloro-2,2-bis(4-chlorophenyl)ethene]

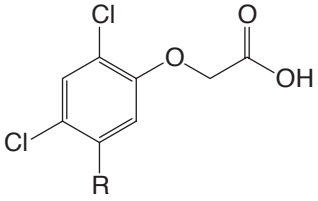
Estimates indicate that nearly 1 billion pounds of DDT were spread throughout the world ecosystem. One pronounced environmental effect of DDE, after conversion from DDT, has been in its action on eggshell formation in many birds. DDE inhibits the enzyme *carbonic anhydrase* that controls the calcium supply for shell formation. As a consequence, the shells are often very fragile and do not survive to the time of hatching. During the late 1940s the populations of eagles, falcons, and hawks dropped dramatically. There can be little doubt that DDT was primarily responsible. DDE also accumulates in the fatty tissues of humans. Although humans appear to have a short-range tolerance to moderate DDE levels, the long-range effects are uncertain.

Study Problem 1 The mechanism for the formation of DDT from chlorobenzene and trichloroacetaldehyde in sulfuric acid involves two electrophilic aromatic substitution reactions. In the first electrophilic substitution reaction, the electrophile is protonated trichloroacetaldehyde. In the second, the electrophile is a carbocation. Propose a mechanism for the formation of DDT.

Study Problem 2 What kind of reaction is involved in the conversion of DDT to DDE?

Organic Halides as Herbicides

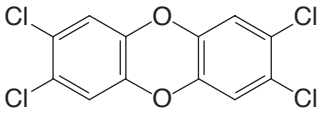
Other chlorinated organic compounds have been used extensively as herbicides. The following two examples are 2,4-D and 2,4,5-T.



R = H; 2,4-D (2,4-dichlorophenoxyacetic acid)

R = Cl; 2,4,5-T (2,4,5-trichlorophenoxyacetic acid)

Enormous quantities of these two compounds were used in an approximately 1:1 mixture as the defoliant Agent Orange during the Vietnam War. Some samples of 2,4,5-T were shown to be teratogenic (a fetus-deforming agent), and its use has been banned in the United States.

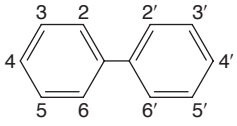


2,3,7,8-Tetrachlorodibenzodioxin (also called TCDD)

This dioxin is also highly stable; it persists in the environment and because of its fat solubility can be passed up the food chain. In sublethal amounts it can cause a disfiguring skin disease called chloracne.

Polychlorinated Biphenyls (PCBs)

Mixtures of polychlorinated biphenyls have been produced and used commercially since 1929. In these mixtures, biphenyls with chlorine atoms at any of the numbered positions (see the following structure) may be present. In all, there are 210 possible compounds. A typical commercial mixture may contain as many as 50 different PCBs. Mixtures are usually classified on the basis of their chlorine content, and most industrial mixtures contain from 40 to 60% chlorine.



Biphenyl

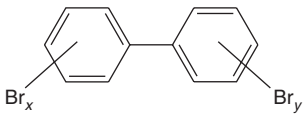
Polychlorinated biphenyls have had a multitude of uses: as heat-exchange agents in transformers; in capacitors, thermostats, and hydraulic systems; as plasticizers in polystyrene coffee cups, frozen food bags, bread wrappers, and plastic liners for baby bottles. They have been used in printing inks, in carbonless carbon paper, and as waxes for making molds for metal castings. Between 1929 and 1972, about 500,000 metric tons of PCBs were manufactured.

Polychlorinated biphenyls are highly persistent in the environment, and, being fat soluble, tend to accumulate in the food chain. PCBs have been found in rainwater, in many species of fish, birds, and other animals (including polar bears) all over the globe, and in human tissue. Fish that feed in PCB-contaminated waters, for example, have PCB levels 1000–100,000 times the level of the surrounding water, and this amount is further magnified in birds that feed on the fish. The toxicity of PCBs depends on the composition of the individual mixture.

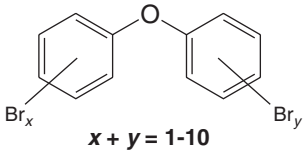
As late as 1975, industrial concerns were legally discharging PCBs into the Hudson River. In 1977, the EPA banned the direct discharge into waterways, and since 1979 their manufacture, processing, and distribution have been prohibited. In 2000 the EPA specified certain sections of the Hudson River for cleanup of PCBs. In 2009, a plan to decontaminate parts of the Hudson River by dredging was finally implemented. See "The Chemistry of . . . Bacterial Dehalogenation of a PCB Derivative" (Section 21.11B) for a potential method of PCB remediation.

Polybrominated Biphenyls and Biphenyl Ethers (PBBs and PBDEs)

As with polychlorinated biphenyls (PCBs), polybrominated aromatic compounds have been used in industry since the early twentieth century. The fire retardant properties of polybrominated and polychlorinated biphenyls and biphenyl ethers, for example, led to their use in building materials, furniture, clothing, and other consumer items. However, the 1970s discovery in Michigan of polybrominated biphenyls (PBBs) in feed for livestock, and subsequently in meat and dairy products, led to suspension of the use of PBBs in 1979.

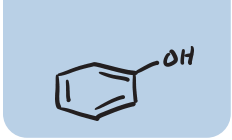


***x* + *y* = 1-10**
Polybrominated biphenyls (PBBs)



***x* + *y* = 1-10**
Polybrominated diphenyl ethers (PBDEs)

(*x* and *y* indicate the possibility of multiple bromine substitution sites on each ring.)



Meanwhile, there is mounting concern about polybromodiphenyl ethers (PBDEs). Although use of PBDEs could potentially save lives and property in their roles as flame retardants, these compounds are now widespread in the environment, and studies have led to significant concern about their toxicity to humans and other animals. As with PCBs, polybrominated biphenyls and polybrominated diphenyl ethers persist in the environment and accumulate in fatty biological tissues. PBDEs have been found in birds, fish, and breast milk. They are now banned in a number of areas.

Key Terms and Concepts

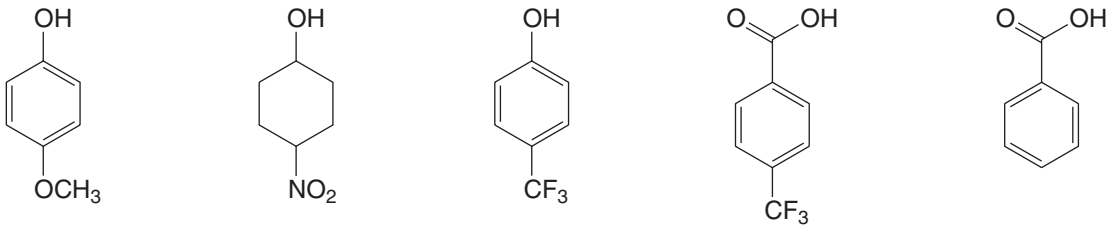
The key terms and concepts that are highlighted in **bold, blue text** within the chapter are defined in the glossary (at the back of the book) and have hyperlinked definitions in the accompanying ***WileyPLUS*** course (www.wileyplus.com).

Problems

Note to Instructors: Many of the homework problems are available for assignment via *WileyPLUS*, an online teaching and learning solution.

PHYSICAL PROPERTIES

21.13 Rank the following in order of increasing acidity.



21.14 Without consulting tables, select the stronger acid from each of the following pairs:

(a) 4-Methylphenol and 4-fluorophenol

(b) 4-Methylphenol and 4-nitrophenol

(c) 4-Nitrophenol and 3-nitrophenol

(d) 4-Methylphenol and benzyl alcohol

(e) 4-Fluorophenol and 4-bromophenol

21.15 What products would be obtained from each of the following acid–base reactions?

(a) Sodium ethoxide in ethanol + phenol $\longrightarrow$

(b) Phenol + aqueous sodium hydroxide $\longrightarrow$

(c) Sodium phenoxide + aqueous hydrochloric acid $\longrightarrow$

(d) Sodium phenoxide + H_2O + CO_2 $\longrightarrow$

21.16 Describe a simple chemical test that could be used to distinguish between members of each of the following pairs of compounds:

(a) 4-Chlorophenol and 4-chloro-1-methylbenzene

(b) 4-Methylphenol and 4-methylbenzoic acid

(c) 4-Methylphenol and 2,4,6-trinitrophenol

(d) Ethyl phenyl ether and 4-ethylphenol

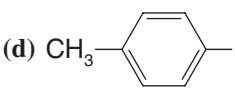
GENERAL REACTIONS

21.17 Complete the following equations:

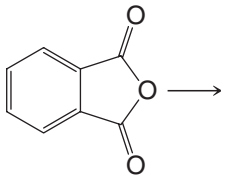
(a) Phenol + Br_2 $\xrightarrow[5°C,\ CS_2]{}$

(b) Phenol + concd H_2SO_4 $\xrightarrow[25°C]{}$

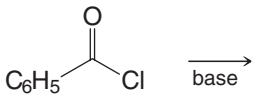
(c) Phenol + concd H_2SO_4 $\xrightarrow[100°C]{}$

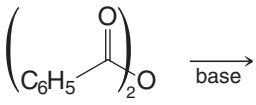
(d)  + *p*-toluenesulfonyl chloride $\xrightarrow[OH^-]{}$

(e) Phenol + Br_2 $\xrightarrow[H_2O]{}$

(f) Phenol + 

(g) *p*-Cresol + Br_2 $\xrightarrow[H_2O]{}$

(h) Phenol + 

(i) Phenol + 

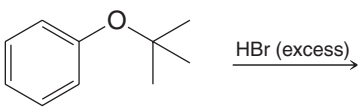
(j) Phenol + NaH ⟶

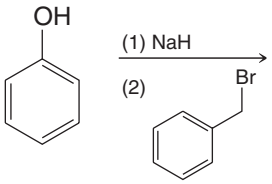
(k) Product of (j) + $CH_3OSO_2OCH_3$ ⟶

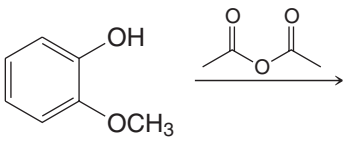
(l) Product of (j) + CH_3I ⟶

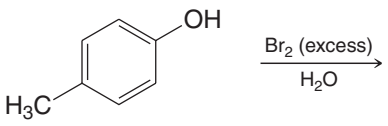
(m) Product of (j) + $C_6H_5CH_2Cl$ ⟶

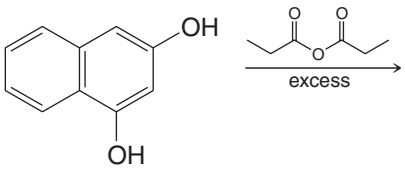
21.18 Predict the product of the following reactions.

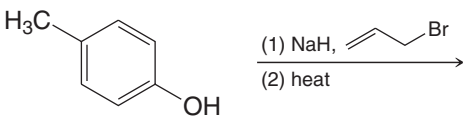
(a) 

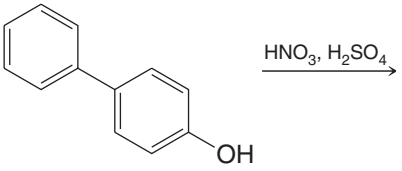
(b) 

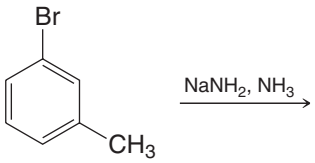
(c) 

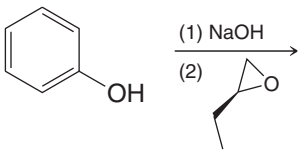
(d) 

(e) 

(f) 

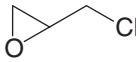
(g) 

(h) 

(i) 

MECHANISMS AND SYNTHESIS

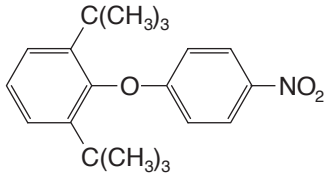
21.19 A synthesis of the β-receptor blocker called toliprolol begins with a reaction between 3-methylphenol and epichlorohydrin. The synthesis is outlined below. Give the structures of the intermediates and of toliprolol.

3-Methylphenol +  ⟶ $C_{10}H_{13}O_2Cl$ $\xrightarrow{OH^-}$

Epichlorohydrin

$C_{10}H_{12}O_2$ $\xrightarrow{(CH_3)_2CHNH_2}$ toliprolol, $C_{13}H_{21}NO_2$

21.20 *p*-Chloronitrobenzene was allowed to react with sodium 2,6-di-*tert*-butylphenoxide with the intention of preparing the diphenyl ether **1**. The product was not **1**, but rather was an isomer of **1** that still possessed a phenolic hydroxyl group.

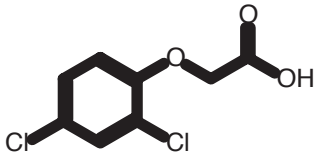


1

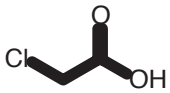
What was this product, and how can one account for its formation?

21.21 When *m*-chlorotoluene is treated with sodium amide in liquid ammonia, the products of the reaction are *o*-, *m*-, and *p*-toluidine (i.e., *o*-$CH_3C_6H_4NH_2$, *m*-$CH_3C_6H_4NH_2$, and *p*-$CH_3C_6H_4NH_2$). Propose plausible mechanisms that account for the formation of each product.

21.22 The herbicide **2,4-D** can be synthesized from phenol and chloroacetic acid. Outline the steps involved.



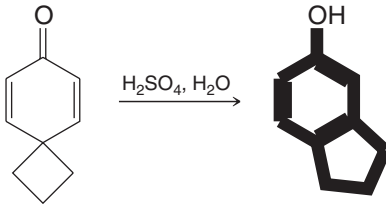
2,4-D
(2,4-dichlorophenoxyacetic acid)



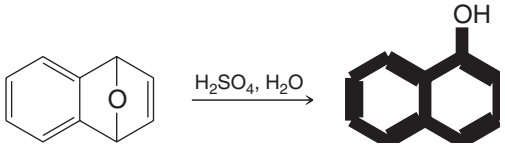
Chloroacetic acid

21.23 The first synthesis of a crown ether (Section 11.16) by C. J. Pedersen (of the DuPont Company) involved treating 1,2-benzenediol with di(2-chloroethyl) ether, $(ClCH_2CH_2)_2O$, in the presence of NaOH. The product was a compound called dibenzo-18-crown-6. Give the structure of dibenzo-18-crown-6 and provide a plausible mechanism for its formation.

21.24 Provide a mechanism for the following reaction.

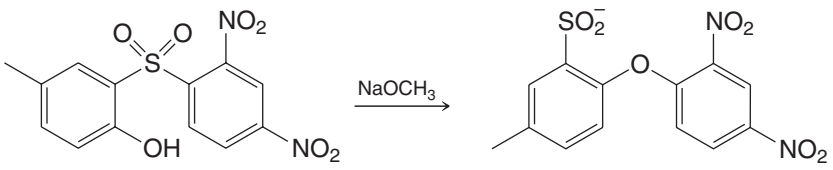


21.25 Provide a mechanism for the following reaction.

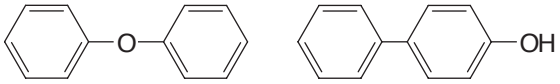


21.26 The widely used antioxidant and food preservative called **BHA** (**b**utylated **h**ydroxy**a**nisole) is actually a mixture of 2-*tert*-butyl-4-methoxyphenol and 3-*tert*-butyl-4-methoxyphenol. BHA is synthesized from *p*-methoxyphenol and 2-methylpropene. **(a)** Suggest how this is done. **(b)** Another widely used antioxidant is BHT (**b**utylated **h**ydroxy**t**oluene). BHT is actually 2,6-di-*tert*-butyl-4-methylphenol, and the raw materials used in its production are *p*-cresol and 2-methylpropene. What reaction is used here?

21.27 Provide a mechanism for the following reaction.

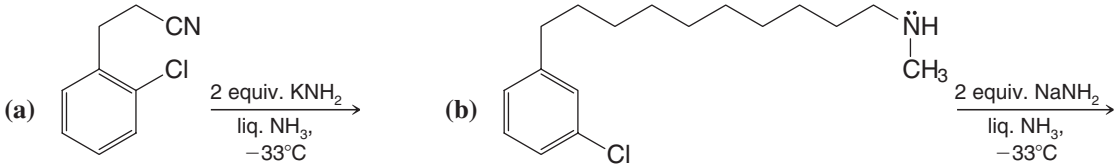


21.28 Account for the fact that the Dow process for the production of phenol produces both diphenyl ether **(1)** and 4-hydroxybiphenyl **(2)** as by-products:

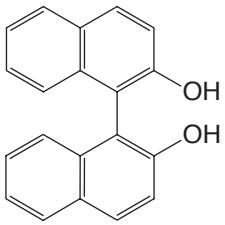


1 **2**

21.29 Predict the outcome of the following reactions:



21.30 Explain how it is possible for 2,2′-dihydroxy-1,1′-binaphthyl (shown at right) to exist in enantiomeric forms.



21.31 Phenols are often effective antioxidants (see Problem 21.26 and "The Chemistry of . . . Antioxidants" in Section 10.11) because they are said to "trap" radicals. The trapping occurs when phenols react with highly reactive radicals to produce less reactive (more stable) phenolic radicals. **(a)** Show how phenol itself might react with an alkoxyl radical (RO·) in a hydrogen abstraction reaction involving the phenolic —OH. **(b)** Write resonance structures for the resulting radical that account for its being relatively unreactive.

SPECTROSCOPY

21.32 A compound **X** ($C_{10}H_{14}O$) dissolves in aqueous sodium hydroxide but is insoluble in aqueous sodium bicarbonate. Compound **X** reacts with bromine in water to yield a dibromo derivative, $C_{10}H_{12}Br_2O$. The 3000–4000 cm^{-1} region of the IR spectrum of **X** shows a broad peak centered at 3250 cm^{-1}; the 680–840 cm^{-1} region shows a strong peak at 830 cm^{-1}. The ^1H NMR spectrum of **X** gives the following: singlet at δ 1.3 (9H), singlet at δ 4.9 (1H), and multiplet at δ 7.0 (4H). What is the structure of **X**?

21.33 Compound **Z** ($C_5H_{10}O$) decolorizes bromine in carbon tetrachloride. The IR spectrum of **Z** shows a broad peak in the 3200–3600 cm^{-1} region. The 300-MHz ^1H NMR spectrum of **Z** is given in Fig. 21.3. Propose a structure for **Z**.

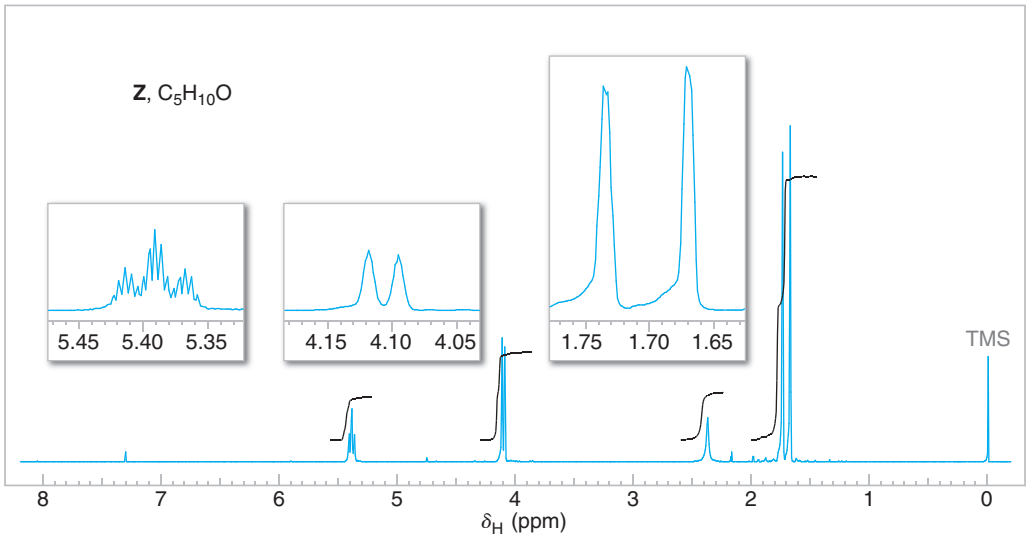
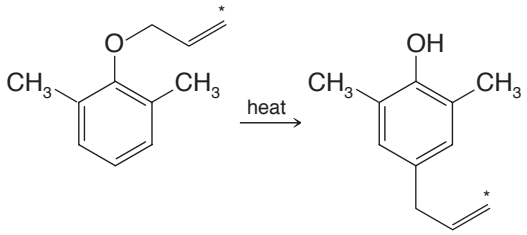


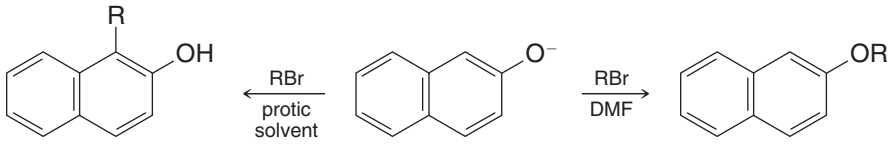
Figure 21.3 The 300-MHz ^1H NMR spectrum of compound **Z** (Problem 21.33). Expansions of the signals are shown in the offset plots.

Challenge Problems

21.34 Explain why, in the case shown, the allyl group has migrated with no change having occurred in the position of the labeled carbon atom within the allyl group:



21.35 In protic solvents the naphthoxide ion (**I**) is alkylated primarily at position 1 (*C*-alkylation) whereas in polar aprotic solvents, such as DMF, the product is almost exclusively the result of a conventional Williamson ether synthesis (*O*-alkylation):

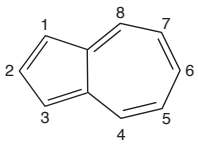


I

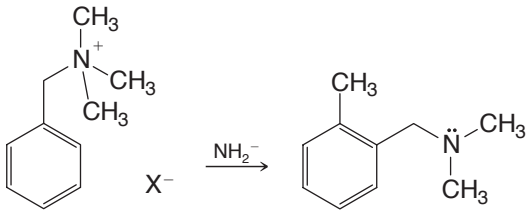
Why does the change in solvent make a difference?

21.36 In comparing nucleophilic aromatic substitution reactions that differ only in the identity of the halogen that is the leaving group in the substrate, it is found that the fluorinated substrate reacts faster than either of the cases where bromine or chlorine is the leaving group. Explain this behavior, which is contrary to the trend among the halogens as leaving groups in S_N1 and S_N2 reactions (in protic solvents).

21.37 In the case of halogen-substituted azulenes, a halogen atom on C6 can be displaced by nucleophiles while one on C1 is unreactive toward nucleophiles. Rationalize this difference in behavior.

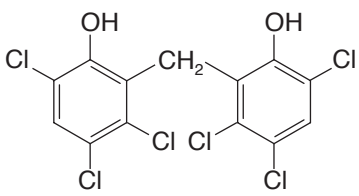


21.38 In the Sommelet–Hauser rearrangement, a benzyl quaternary ammonium salt reacts with a strong base to give a benzyl tertiary amine, as exemplified below:

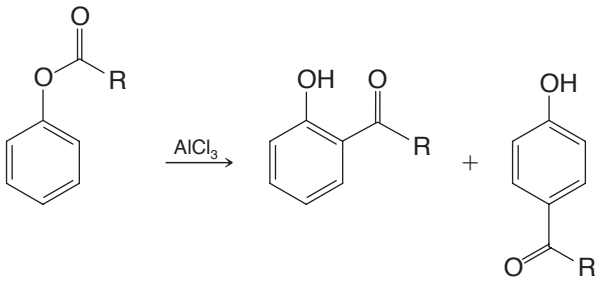


Suggest a mechanism for this rearrangement.

21.39 Hexachlorophene was a widely used germicide until it was banned in 1972 after tests showed that it caused brain damage in test animals. Suggest how this compound might be synthesized, starting with benzene.



21.40 The Fries rearrangement occurs when a phenolic ester is heated with a Friedel–Crafts catalyst such as $AlCl_3$:



The reaction may produce both ortho and para acylated phenols, the former generally favored by high temperatures and the latter by low temperatures. **(a)** Suggest an experiment that might indicate whether the reaction is inter- or intramolecular. **(b)** Explain the temperature effect on product formation.

21.41 Compound **W** was isolated from a marine annelid commonly used in Japan as a fish bait, and it was shown to be the substance that gives this organism its observed toxicity to some insects that contact it.

MS (*m/z*): 151 (relative abundance 0.09), 149 ($M\cdot^+$, rel. abund. 1.00), 148

IR (cm^{-1}): 2960, 2850, 2775

^{1}H NMR (δ): 2.3 (s, 6H), 2.6 (d, 4H), and 3.2 (pentet, 1H)

^{13}C NMR (δ): 38 (CH_3), 43 (CH_2), and 75 (CH)

These reactions were used to obtain further information about the structure of **W**:

$$\mathbf{W} \xrightarrow{NaBH_4} \mathbf{X} \xrightarrow{C_6H_5COCl} \mathbf{Y} \xrightarrow{\text{Raney Ni}} \mathbf{Z}$$

Compound **X** had a new infrared band at 2570 cm^{-1} and these NMR data:

^{1}H NMR (δ): 1.6 (t, 2H), 2.3 (s, 6H), 2.6 (m, 4H), and 3.2 (pentet, 1H)

^{13}C NMR (δ): 28 (CH_2), 38 (CH_3), and 70 (CH)

Compound **Y** had these data:

IR (cm^{-1}): 3050, 2960, 2850, 1700, 1610, 1500, 760, 690

^{1}H NMR (δ): 2.3 (s, 6H), 2.9 (d, 4H), 3.0 (pentet, 1H), 7.4 (m, 4H), 7.6 (m, 2H), and 8.0 (m, 4H)

^{13}C NMR (δ): 34 (CH_2), 39 (CH_3), 61 (CH), 128 (CH), 129 (CH), 134 (CH), 135 (C), and 187 (C)

Compound **Z** had

MS (*m/z*): 87 ($M\cdot^+$), 86, 72

IR (cm^{-1}): 2960, 2850, 1385, 1370, 1170

^{1}H NMR (δ): 1.0 (d, 6H), 2.3 (s, 6H), and 3.0 (heptet, 1H)

^{13}C NMR (δ): 21 (CH_3), 39 (CH_3), and 55 (CH)

What are the structures of **W** and of each of its reaction products **X, Y**, and **Z**?

21.42 Phenols generally are not changed on treatment with sodium borohydride followed by acidification to destroy the excess, unreacted hydride. For example, the 1,2-, 1,3-, and 1,4-benzenediols and 1,2,3-benzenetriol are unchanged under these conditions. However, 1,3,5-benzenetriol (phloroglucinol) gives a high yield of a product **A** that has these properties:

MS (*m/z*): 110

IR (cm^{-1}): 3250 (broad), 1613, 1485

^{1}H NMR (δ in DMSO): 6.15 (m, 3H), 6.89 (t, 1H), and 9.12 (s, 2H)

(a) What is the structure of **A**?

(b) Suggest a mechanism by which the above reaction occurred. (1,3,5-Benzenetriol is known to have more tendency to exist in a keto tautomeric form than do simpler phenols.)

21.43 Open the molecular model file for benzyne and examine the following molecular orbitals: the LUMO (lowest unoccupied molecular orbital), the HOMO (highest occupied molecular orbital), the HOMO-1 (next lower energy orbital), the HOMO-2 (next lower in energy), and the HOMO-3 (next lower in energy). **(a)** Which orbital best represents the region where electrons of the additional π bond in benzyne would be found? **(b)** Which orbital would accept electrons from a Lewis base on nucleophilic addition to benzyne? **(c)** Which orbitals are associated with the six π electrons of the aromatic system? Recall that each molecular orbital can hold a maximum of two electrons.

Learning Group Problems

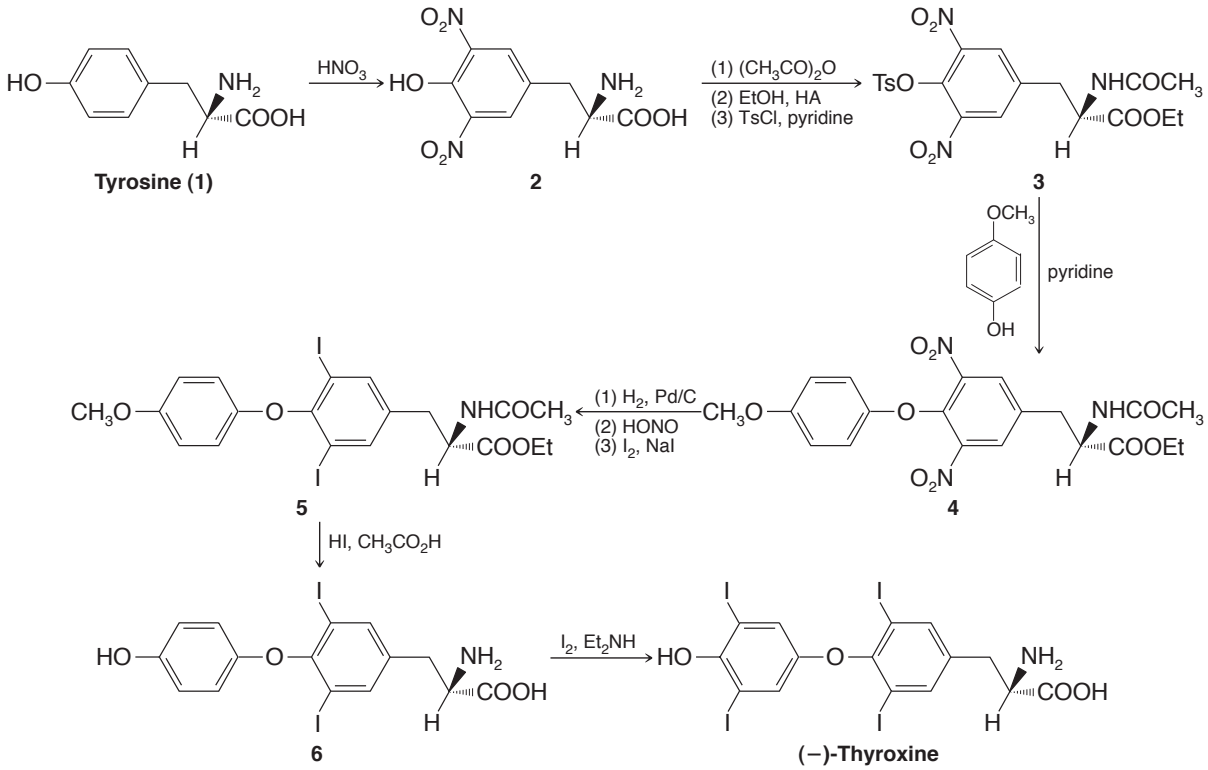
1. Thyroxine is a hormone produced by the thyroid gland that is involved in regulating metabolic activity. In a previous Learning Group Problem (Chapter 15) we considered reactions involved in a chemical synthesis of thyroxine. The following is a synthesis of optically pure thyroxine from the amino acid tyrosine (also see Problem 2, below). This synthesis proved to be useful on an industrial scale. (Scheme adapted from Fleming, I., *Selected Organic Syntheses*, pp. 31–33. © 1973 John Wiley & Sons, Limited. Reproduced with permission.)

(a) 1 to **2** What type of reaction is involved in the conversion of **1** to **2**? Write a detailed mechanism for this transformation. Explain why the nitro groups appear where they do in **2**.

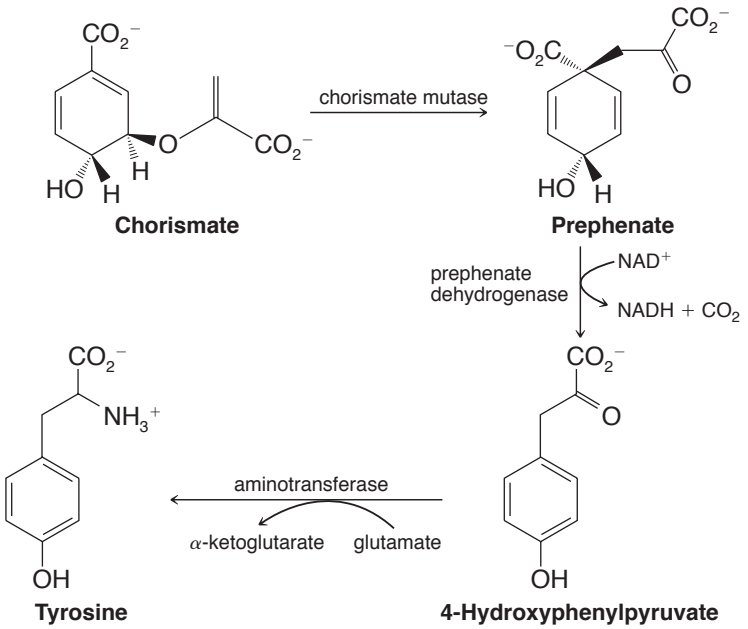
(b) 2 to **3** (i) Write a detailed mechanism for step (1) in the conversion of **2** to **3**.
(ii) Write a detailed mechanism for step (2) in the conversion of **2** to **3**.
(iii) Write a detailed mechanism for step (3) in the conversion of **2** to **3**.

(c) 3 to **4** (i) What type of reaction mechanism is involved in the conversion of **3** to **4**?
(ii) Write a detailed mechanism for the reaction from **3** to **4**. What key intermediate is involved?

(d) 5 to **6** Write a detailed mechanism for conversion of the methoxyl group of **5** to the phenolic hydroxyl of **6**.

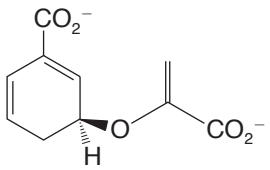


2. Tyrosine is an amino acid with a phenolic side chain. Biosynthesis in plants and microbes of tyrosine involves enzymatic conversion of chorismate to prephenate, below. Prephenate is then processed further to form tyrosine. These steps are shown here:



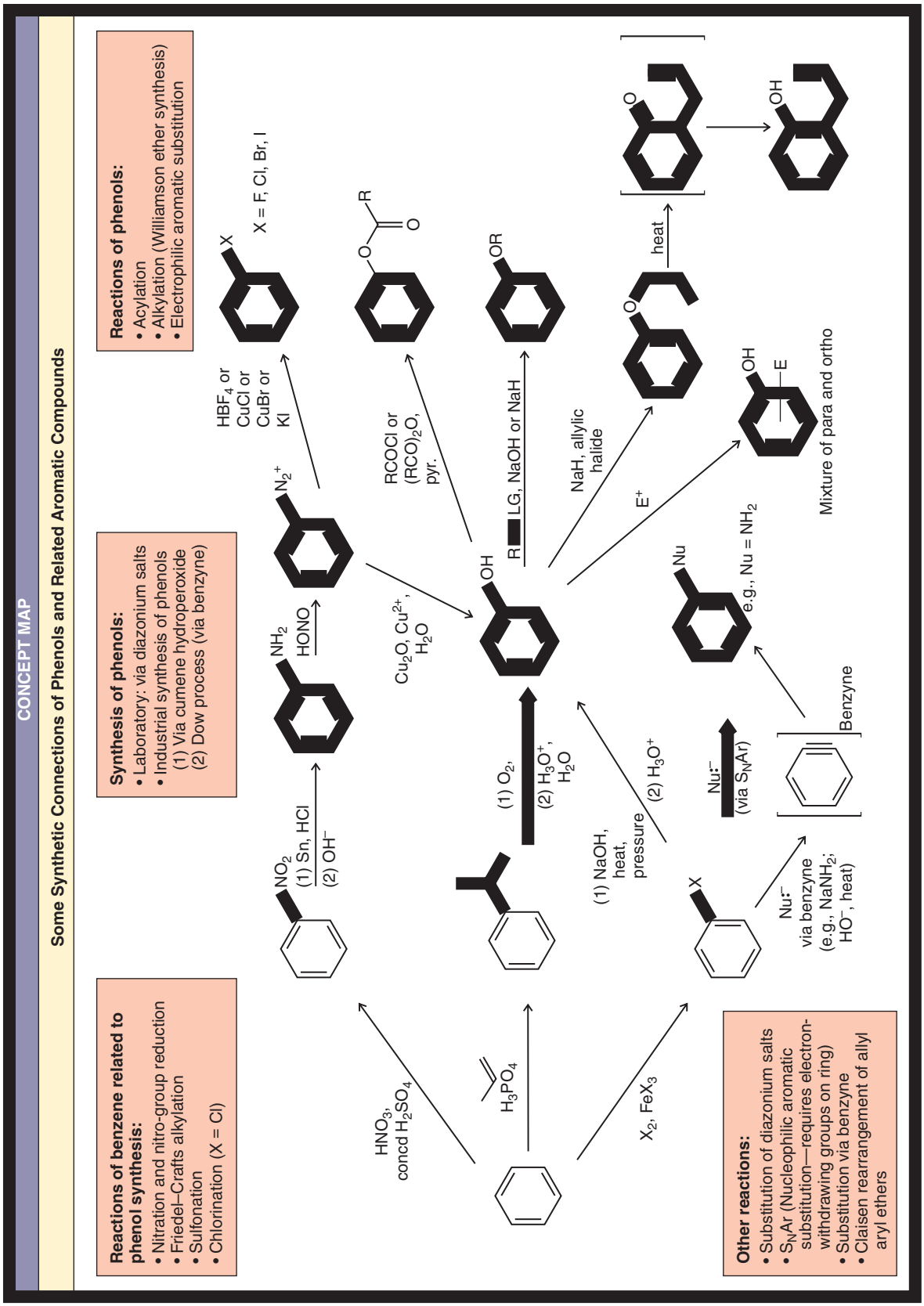
(a) There has been substantial research and debate about the enzymatic conversion of chorismate to prephenate by chorismate mutase. Although the enzymatic mechanism may not be precisely analogous, what laboratory reaction have we studied in this chapter that resembles the biochemical conversion of chorismate to prephenate? Draw arrows to show the movement of electrons involved in such a reaction from chorismate to prephenate.

(b) When the type of reaction you proposed above is applied in laboratory syntheses, it is generally the case that the reaction proceeds by a concerted chair conformation transition state. Five of the atoms of the chair are carbon and one is oxygen. In both the reactant and product, the chair has one bond missing, but at the point of the bond reorganization there is roughly concerted flow of electron density throughout the atoms involved in the chair. For the reactant shown below, draw the structure of the product and the associated chair conformation transition state for this type of reaction:



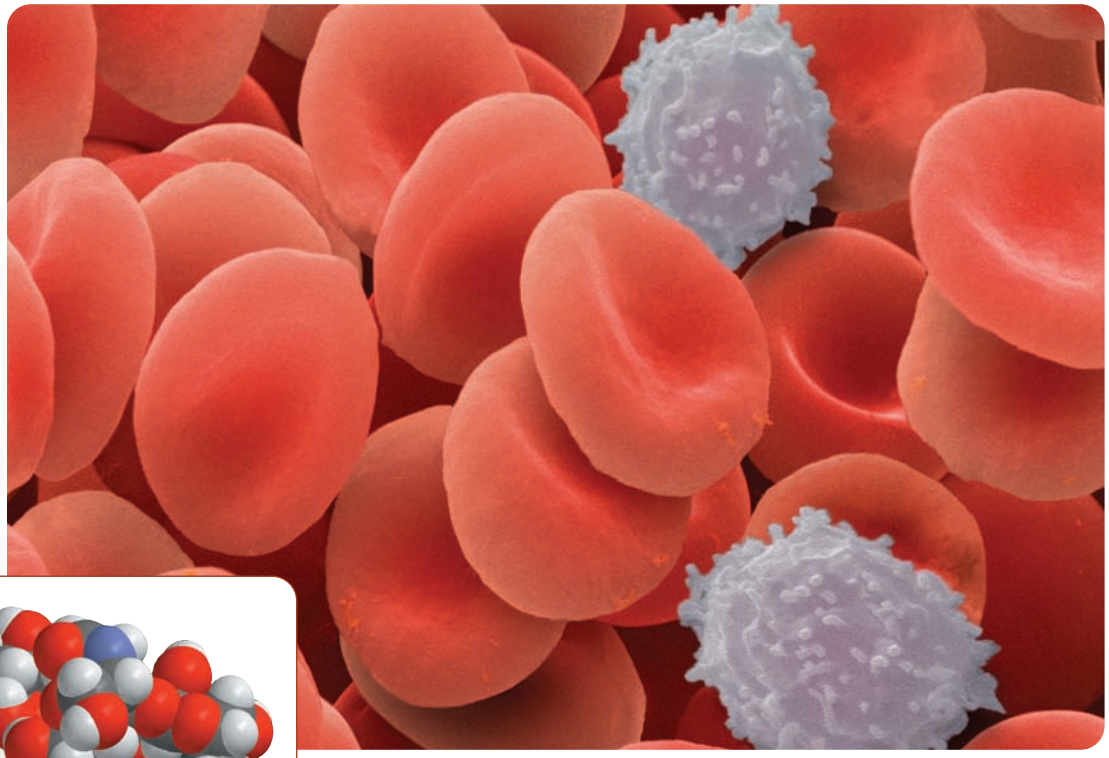
(c) Draw the structure of the nicotinamide ring of NAD^+ and draw mechanism arrows to show the decarboxylation of prephenate to 4-hydroxyphenylpyruvate with transfer of the hydride to NAD^+ (this is the type of process involved in the mechanism of prephenate dehydrogenase).

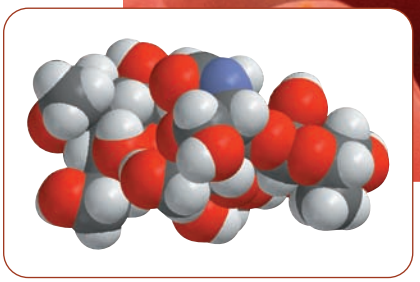
(d) Look up the structures of glutamate (glutamic acid) and α-ketoglutarate and consider the process of transamination involved in conversion of 4-hydroxyphenylpyruvate to tyrosine. Identify the source of the amino group in this transamination (i.e., what is the amino group "donor"?). What functional group is left after the amino group has been transferred from its donor? Propose a mechanism for this transamination. Note that the mechanism you propose will likely involve formation and hydrolysis of several imine intermediates—reactions similar to others we studied in Section 16.8.



22

Carbohydrates



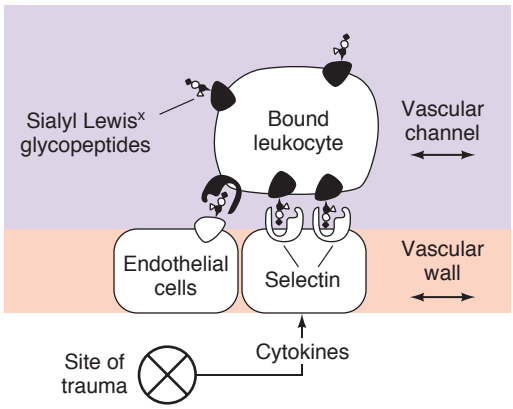


Sialyl Lewis^x, a carbohydrate that is important in the recognition and healing of traumatized tissue.

White blood cells continually patrol the circulatory system and interstitial spaces, ready for mobilization at a site of trauma. The frontline scouts for leukocytes are carbohydrate groups on their surface called sialyl Lewis^x acids. When injury occurs, cells at the site of trauma display proteins, called selectins, that signal the site of injury and bind sialyl Lewis^x acids. Binding between selectins and the sialyl Lewis^x acids on the leukocytes causes adhesion of leukocytes at the affected area. Recruitment of leukocytes in this way is an important step in the inflammatory cascade. It is a necessary part of the healing process as well as part of our natural defense against infection. A molecular model of a sialyl Lewis^x acid is shown above, and its structural formula is given in Section 22.16.

There are some maladies, however, that result from the over-enthusiastic recruitment of leukocytes. Rheumatoid arthritis, strokes, and injuries related to perfusion during surgery and organ transplantation are a few examples. In these conditions, the body perceives that certain cells are under duress, and it reacts accordingly to initiate the inflammatory cascade. Unfortunately, under these circumstances the inflammatory cascade actually causes greater harm than good.

A strategy for combating undesirable initiation of the inflammatory cascade is to disrupt the adhesion of leukocytes. This can be done by blocking the selectin binding sites for sialyl Lewis^x acids. Chemists have advanced this approach by synthesizing both natural and mimetic sialyl Lewis^x acids for studies on the binding process. These compounds have helped identify key functional groups in sialyl Lewis^x acids that are required



Patrolling leukocytes bind at the site of trauma by interactions between sialyl Lewis^x glycoproteins on their surface and selectin proteins on the injured cell. (Reprinted with permission from Simanek, E.E.; McGarvey, G.J.; Jablonowski, J.A.; Wong, C.A., *Chemical Reviews*, *98*, p. 835, Figure 1, 1998. Copyright 1998 American Chemical Society.)

for recognition and binding. Chemists have even designed and synthesized novel compounds that have tighter binding affinities than the natural sialyl Lewis^x acids. Among them are polymers with repeating occurrences of the structural motifs essential for binding. These polymeric species presumably occupy multiple sialyl Lewis^x acid binding sites at once, thereby binding more tightly than monomeric sialyl Lewis^x acid analogs.

Efforts like these to prepare finely tuned molecular agents are typical of research in drug discovery and design. In the case of sialyl Lewis^x acid analogs, chemists hope to create new therapies for chronic inflammatory diseases by making ever-improved agents for blocking undesired leukocyte adhesion.

22.1 Introduction

22.1A Classification of Carbohydrates

The group of compounds known as carbohydrates received their general name because of early observations that they often have the formula $C_x(H_2O)_y$—that is, they appear to be "hydrates of carbon." Simple carbohydrates are also known as sugars or saccharides (Latin *saccharum*, Greek *sakcharon*, sugar) and the ending of the names of most sugars is *-ose*. Thus, we have such names as *sucrose* for ordinary table sugar, *glucose* for the principal sugar in blood, *fructose* for a sugar in fruits and honey, and *maltose* for malt sugar.

Helpful Hint

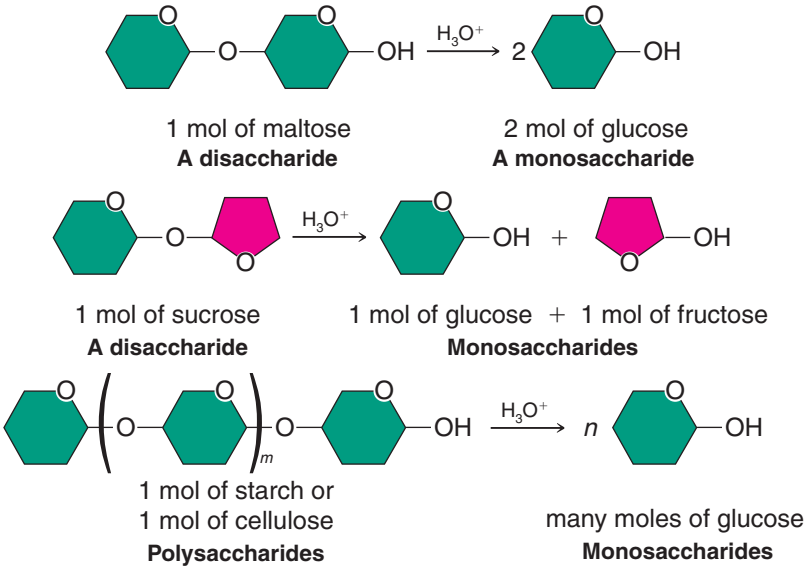
You may find it helpful now to review the chemistry of hemiacetals and acetals (Section 16.7).

- **Carbohydrates** are usually defined as polyhydroxy aldehydes and ketones or substances that hydrolyze to yield polyhydroxy aldehydes and ketones. They exist primarily in their hemiacetal or acetal forms (Section 16.7).

The simplest carbohydrates, those that cannot be hydrolyzed into simpler carbohydrates, are called **monosaccharides**. On a molecular basis, carbohydrates that undergo hydrolysis to produce only 2 molecules of monosaccharide are called **disaccharides**; those that yield 3 molecules of monosaccharide are called **trisaccharides**; and so on. (Carbohydrates that hydrolyze to yield 2–10 molecules of monosaccharide are sometimes called **oligosaccharides**.) Carbohydrates that yield a large number of molecules of monosaccharides (>10) are known as **polysaccharides**.

Maltose and sucrose are examples of disaccharides. On hydrolysis, 1 mol of maltose yields 2 mol of the monosaccharide glucose; sucrose undergoes hydrolysis to yield 1 mol of glucose and 1 mol of the monosaccharide fructose. Starch and cellulose are examples

of polysaccharides; both are glucose polymers. Hydrolysis of either yields a large number of glucose units. The following shows these hydrolyses in a schematic way:



Carbohydrates are the most abundant organic constituents of plants. They not only serve as an important source of chemical energy for living organisms (sugars and starches are important in this respect), but also in plants and in some animals they serve as important constituents of supporting tissues (this is the primary function of the cellulose found in wood, cotton, and flax, for example).

We encounter carbohydrates at almost every turn of our daily lives. The paper on which this book is printed is largely cellulose; so, too, is the cotton of our clothes and the wood of our houses. The flour from which we make bread is mainly starch, and starch is also a major constituent of many other foodstuffs, such as potatoes, rice, beans, corn, and peas. Carbohydrates are central to metabolism, and they are important for cell recognition (see the chapter opening vignette and Section 22.16).

22.1B Photosynthesis and Carbohydrate Metabolism

Schematic diagram of a chloroplast from corn. (Reprinted with permission of John Wiley & Sons, Inc., from Voet, D. and Voet, J. G., *Biochemistry*, Second Edition. © 1995 Voet, D. and Voet, J. G.)

Carbohydrates are synthesized in green plants by *photosynthesis*—a process that uses solar energy to reduce, or "fix," carbon dioxide. Photosynthesis in algae and higher plants occurs in cell organelles called chloroplasts. The overall equation for photosynthesis can be written as follows:

$$x\,CO_2 + y\,H_2O + \text{solar energy} \longrightarrow \underset{\text{Carbohydrate}}{C_x(H_2O)_y} + x\,O_2$$

Many individual enzyme-catalyzed reactions take place in the general photosynthetic process and not all are fully understood. We know, however, that photosynthesis begins with the absorption of light by the important green pigment of plants, chlorophyll (Fig. 22.1). The green color of chlorophyll and, therefore, its ability to absorb sunlight in the visible region are due primarily to its extended conjugated system. As photons of sunlight are trapped by chlorophyll, energy becomes available to the plant in a chemical form that can be used to carry out the reactions that reduce carbon dioxide to carbohydrates and oxidize water to oxygen.

Carbohydrates act as a major chemical repository for solar energy. Their energy is released when animals or plants metabolize carbohydrates to carbon dioxide and water:

$$C_x(H_2O)_y + x\,O_2 \longrightarrow x\,CO_2 + y\,H_2O + \text{energy}$$

The metabolism of carbohydrates also takes place through a series of enzyme-catalyzed reactions in which each energy-yielding step is an oxidation (or the consequence of an oxidation).

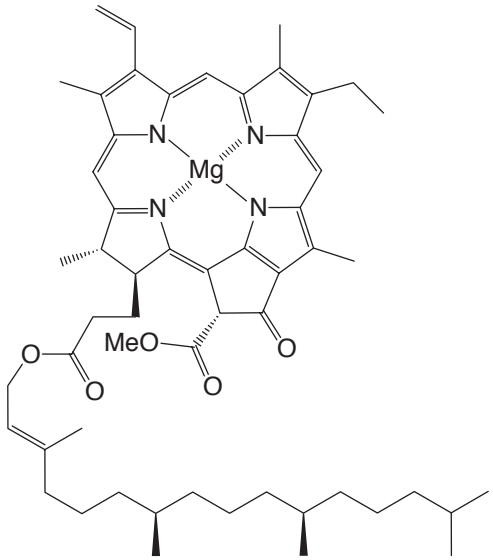


Figure 22.1 Chlorophyll *a*. [The structure of chlorophyll *a* was established largely through the work of H. Fischer (Munich), R. Willstätter (Munich), and J. B. Conant (Harvard). A synthesis of chlorophyll *a* from simple organic compounds was achieved by R. B. Woodward (Harvard) in 1960, who won the Nobel prize in 1965 for his outstanding contributions to synthetic organic chemistry.]

Although some of the energy released in the oxidation of carbohydrates is inevitably converted to heat, much of it is conserved in a new chemical form through reactions that are coupled to the synthesis of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate (P_i) (Fig. 22.2). The phosphoric anhydride bond that forms between the terminal phosphate group of ADP and the phosphate ion becomes another repository of chemical energy. Plants and animals can use the conserved energy of ATP (or very similar substances) to carry out all of their energy-requiring processes: the contraction of a muscle,

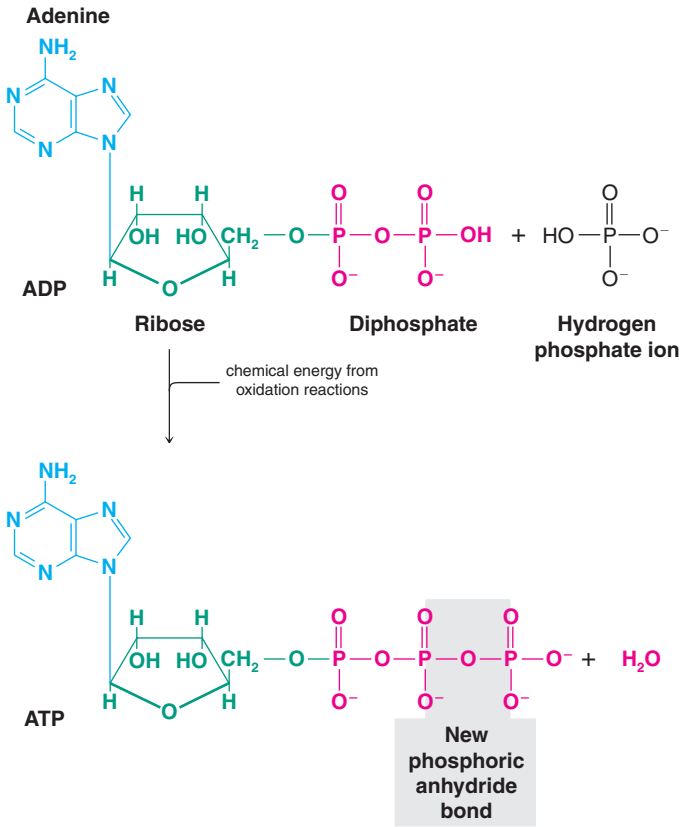
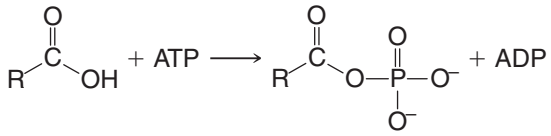


Figure 22.2 The synthesis of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and hydrogen phosphate ion. This reaction takes place in all living organisms, and adenosine triphosphate is the major compound into which the chemical energy released by biological oxidations is transformed.

the synthesis of a macromolecule, and so on. When the energy in ATP is used, a coupled reaction takes place in which ATP is hydrolyzed,

$$ATP + H_2O \longrightarrow ADP + P_i + \text{energy}$$

or a new anhydride linkage is created,

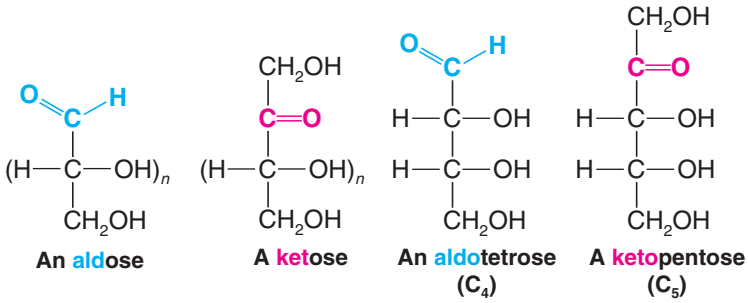


An acyl phosphate

22.2 Monosaccharides

22.2A Classification of Monosaccharides

Monosaccharides are classified according to (1) the number of carbon atoms present in the molecule and (2) whether they contain an aldehyde or keto group. Thus, a monosaccharide containing three carbon atoms is called a *triose*; one containing four carbon atoms is called a *tetrose*; one containing five carbon atoms is a *pentose*; and one containing six carbon atoms is a *hexose*. A monosaccharide containing an aldehyde group is called an **aldose**; one containing a keto group is called a **ketose**. These two classifications are frequently combined. A C_4 aldose, for example, is called an *aldotetrose*; a C_5 ketose is called a *ketopentose*.

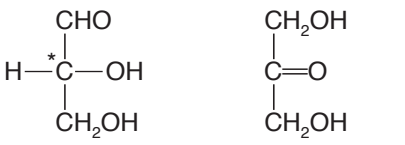


Review Problem 22.1

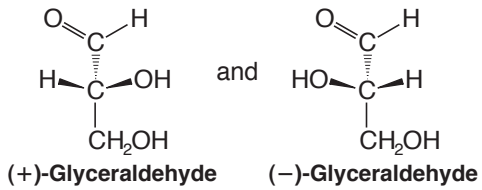
How many chirality centers are contained in **(a)** the aldotetrose and **(b)** the ketopentose just given? **(c)** How many stereoisomers would you expect from each general structure?

22.2B D and L Designations of Monosaccharides

The simplest monosaccharides are the compounds glyceraldehyde and dihydroxyacetone (see the following structures). Of these two compounds, only glyceraldehyde contains a chirality center.



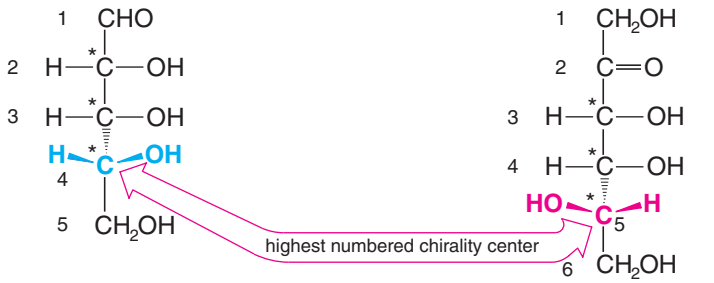
Glyceraldehyde exists, therefore, in two enantiomeric forms that are known to have the absolute configurations shown here:



(+)-Glyceraldehyde **(−)-Glyceraldehyde**

We saw in Section 5.7 that, according to the Cahn–Ingold–Prelog convention, (+)-glyceraldehyde should be designated (*R*)-(+)-glyceraldehyde and (−)-glyceraldehyde should be designated (*S*)-(−)-glyceraldehyde.

Early in the twentieth century, before the absolute configurations of any organic compounds were known, another system of stereochemical designations was introduced. According to this system (first suggested by M. A. Rosanoff of New York University in 1906), (+)-glyceraldehyde is designated D-(+)-glyceraldehyde and (−)-glyceraldehyde is designated L-(−)-glyceraldehyde. These two compounds, moreover, serve as configurational standards for all monosaccharides. A monosaccharide *whose highest numbered chirality center* (the penultimate carbon) has the same configuration as D-(+)-glyceraldehyde is designated as a D sugar; one whose highest numbered chirality center has the same configuration as L-glyceraldehyde is designated as an L sugar. By convention, acyclic forms of monosaccharides are drawn vertically with the aldehyde or keto group at or nearest the top. When drawn in this way, D sugars have the —OH on their penultimate carbon on the right:



A D-aldopentose **An L-ketohexose**

The **D and L nomenclature** designations are like (*R*) and (*S*) designations in that they are not necessarily related to the optical rotations of the sugars to which they are applied. Thus, one may encounter other sugars that are D-(+) or D-(−) and ones that are L-(+) or L-(−).

The D–L system of stereochemical designations is thoroughly entrenched in the literature of carbohydrate chemistry, and even though it has the disadvantage of specifying the configuration of only one chirality center—that of the highest numbered chirality center—we shall employ the D–L system in our designations of carbohydrates.

Review Problem 22.2

Write three-dimensional formulas for each aldotetrose and ketopentose isomer in Review Problem 22.1 and designate each as a D or L sugar.

22.2C Structural Formulas for Monosaccharides

Later in this chapter we shall see how the great carbohydrate chemist Emil Fischer* was able to establish the stereochemical configuration of the aldohexose D-(+)-glucose, the most abundant monosaccharide. In the meantime, however, we can use D-(+)-glucose as an example illustrating the various ways of representing the structures of monosaccharides.

*Emil Fischer (1852–1919) was professor of organic chemistry at the University of Berlin. In addition to monumental work in the field of carbohydrate chemistry, where Fischer and co-workers established the configuration of most of the monosaccharides, Fischer also made important contributions to studies of amino acids, proteins, purines, indoles, and stereochemistry generally. As a graduate student, Fischer discovered phenylhydrazine, a reagent that was highly important in his later work with carbohydrates. Fischer was the second recipient (in 1902) of the Nobel Prize in Chemistry.

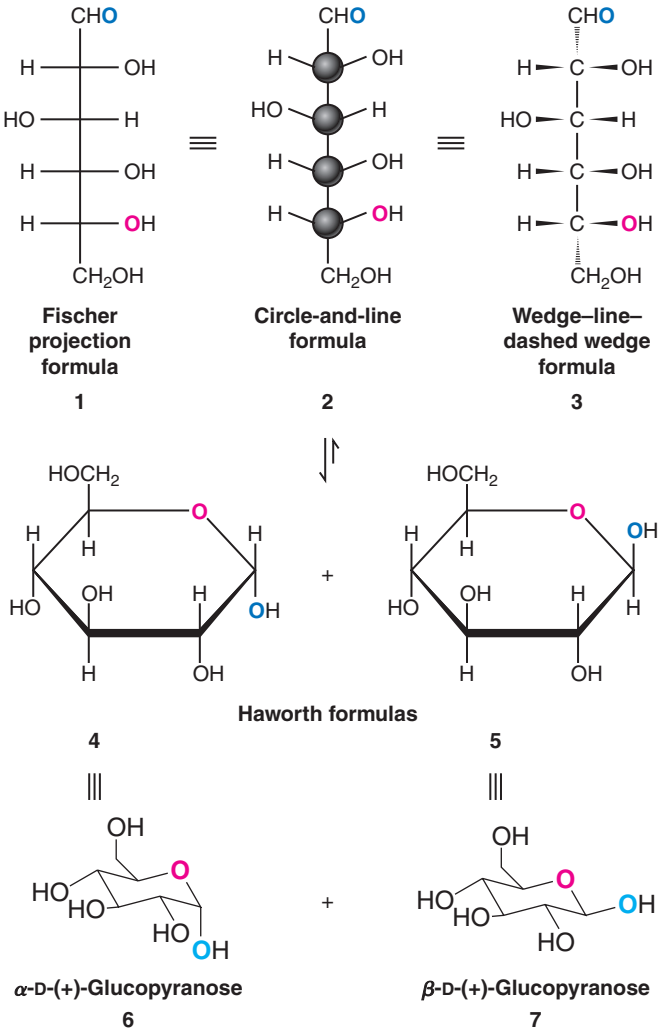


Figure 22.3 Formulas 1–3 are used for the open-chain structure of D-(+)-glucose. Formulas 4–7 are used for the two cyclic hemiacetal forms of D-(+)-glucose.

Fischer represented the structure of D-(+)-glucose with the cross formulation **(1)** in Fig. 22.3. This type of formulation is now called a **Fischer projection** (Section 5.13) and is still useful for carbohydrates. In Fischer projections, by convention, *horizontal lines project out toward the reader and vertical lines project behind the plane of the page. When we use Fischer projections, however, we must not* (in our mind's eye) *remove them from the plane of the page in order to test their superposability and we must not rotate them by 90°*. In terms of more familiar formulations, the Fischer projection translates into formulas **6** and **7**. In IUPAC nomenclature and with the Cahn–Ingold–Prelog system of stereochemical designations, the open-chain form of D-(+)-glucose is (2*R*,3*S*,4*R*,5*R*)-2,3,4,5,6-pentahydroxyhexanal.

Helpful Hint

Use molecular models to help you learn to interpret Fischer projection formulas.

The meaning of formulas **1, 2**, and **3** can be seen best through the use of molecular models: We first construct a chain of six carbon atoms with the —CHO group at the top and a —CH_2OH group at the bottom. We then bring the CH_2OH group up behind the chain until it almost touches the —CHO group. Holding this model so that the —CHO and —CH_2OH groups are directed generally away from us, we then begin placing —H and —OH groups on each of the four remaining carbon atoms. The —OH group of C2 is placed on the right; that of C3 on the left; and those of C4 and C5 on the right.

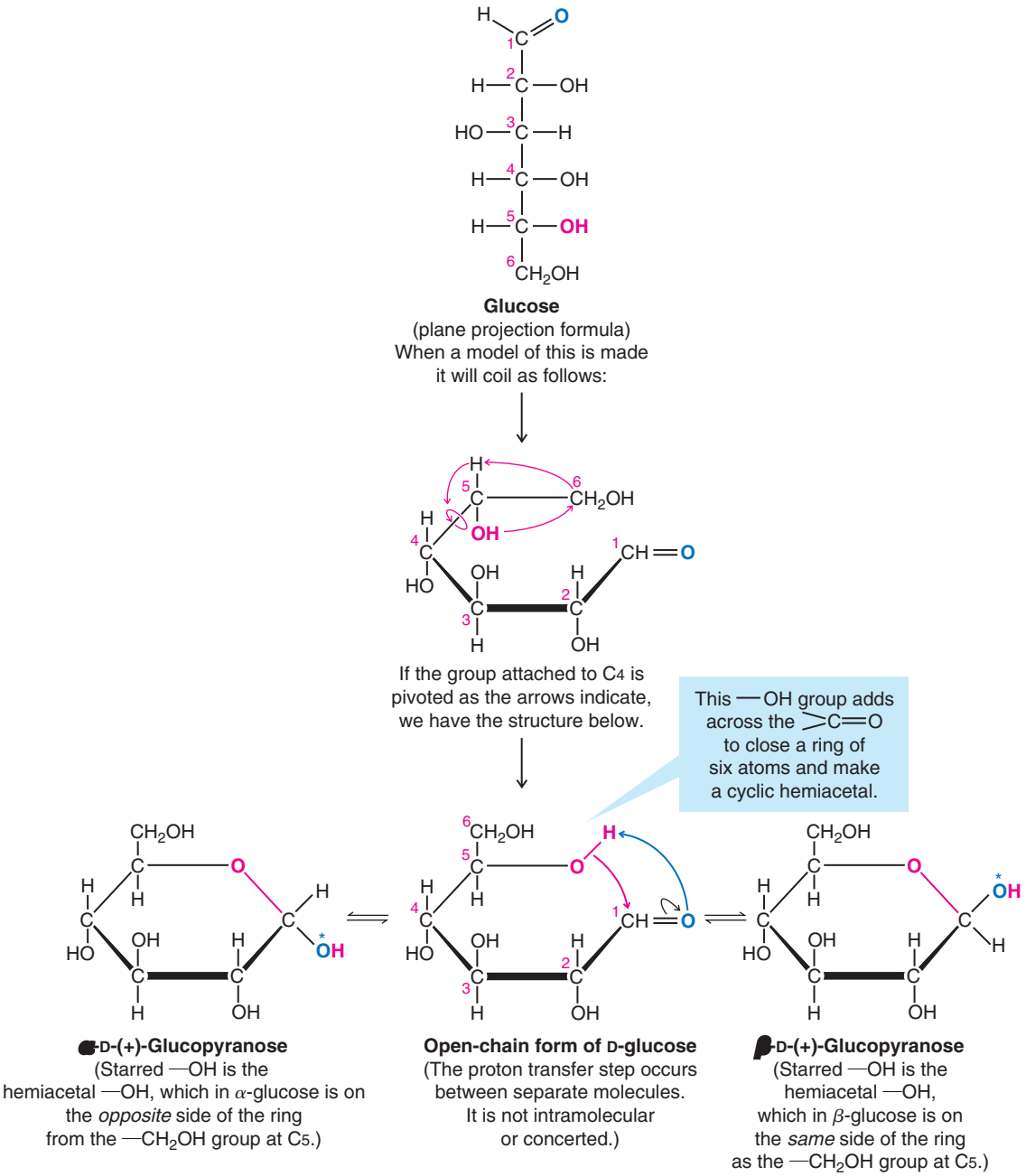


Figure 22.4 Haworth formulas for the cyclic hemiacetal forms of D-(+)-glucose and their relation to the open-chain polyhydroxy aldehyde structure. (Reprinted with permission of John Wiley & Sons, Inc., from Holum, J. R., *Organic Chemistry: A Brief Course*, p. 316. Copyright 1975.)

Although many of the properties of D-(+)-glucose can be explained in terms of an open-chain structure (**1, 2**, or **3**), a considerable body of evidence indicates that the open-chain structure exists, primarily, in equilibrium with two cyclic forms. These can be represented by structures **4** and **5** or **6** and **7**. The cyclic forms of D-(+)-glucose are **hemiacetals** formed by an intramolecular reaction of the —OH group at C5 with the aldehyde group (Fig. 22.4). Cyclization creates a new chirality center at C1, and this chirality center explains how two cyclic forms are possible. These two cyclic forms are *diastereomers* that differ only in the configuration of C1.

- In carbohydrate chemistry diastereomers differing only at the hemiacetal or acetal carbon are called **anomers**, and the hemiacetal or acetal carbon atom is called the **anomeric carbon atom**.

Structures **4** and **5** for the glucose anomers are called **Haworth formulas*** and, although they do not give an accurate picture of the shape of the six-membered ring, they have many practical uses. Figure 22.4 demonstrates how the representation of each chirality center of the open-chain form can be correlated with its representation in the Haworth formula.

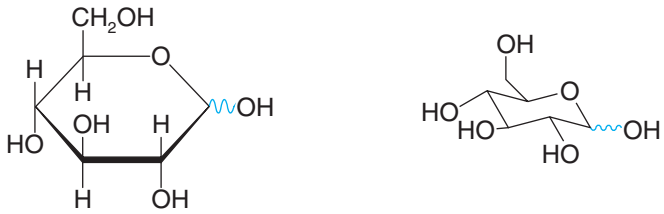
Helpful Hint

α and β also find common use in steroid nomenclature (Section 23.4A).

Each glucose anomer is designated as an **α anomer** or a **β anomer** depending on the location of the —OH group of C1. When we draw the cyclic forms of a D sugar in the orientation shown in Figs. 22.3 or 22.4, the α anomer has the —OH trans to the $—CH_2OH$ group and the β anomer has the —OH cis to the $—CH_2OH$ group.

Studies of the structures of the cyclic hemiacetal forms of D-(+)-glucose using X-ray analysis have demonstrated that the actual conformations of the rings are the chair forms represented by conformational formulas **6** and **7** in Fig. 22.3. This shape is exactly what we would expect from our studies of the conformations of cyclohexane (Chapter 4), and it is especially interesting to notice that in the β anomer of D-glucose all of the large substituents, —OH and $—CH_2OH$, are equatorial. In the α anomer, the only bulky axial substituent is the —OH at C1.

It is convenient at times to represent the cyclic structures of a monosaccharide without specifying whether the configuration of the anomeric carbon atom is α or β. When we do this, we shall use formulas such as the following:



The symbol ∿ indicates α or β (three-dimensional view not specified).

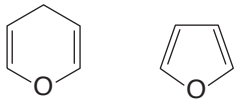
Not all carbohydrates exist in equilibrium with six-membered hemiacetal rings; in several instances the ring is five membered. (Even glucose exists, to a small extent, in equilibrium with five-membered hemiacetal rings.) Because of this variation, a system of nomenclature has been introduced to allow designation of the ring size.

- If the monosaccharide ring is six membered, the compound is called a **pyranose**; if the ring is five membered, the compound is designated as a **furanose**.**

Thus, the full name of compound **4** (or **6**) is α-D-(+)-glucopyranose, while that of **5** (or **7**) is β-D-(+)-glucopyranose.

*Haworth formulas are named after the English chemist W. N. Haworth (University of Birmingham), who, in 1926, along with E. L. Hirst, demonstrated that the cyclic form of glucose acetals consists of a six-membered ring. Haworth received the Nobel prize for his work in carbohydrate chemistry in 1937. For an excellent discussion of Haworth formulas and their relation to open-chain forms, see "The Conversion of Open Chain Structures of Monosaccharides into the Corresponding Haworth Formulas," Wheeler, D. M. S., Wheeler, M. M., and Wheeler, T. S., *J. Chem. Educ*. **1982**, *59*, 969–970.

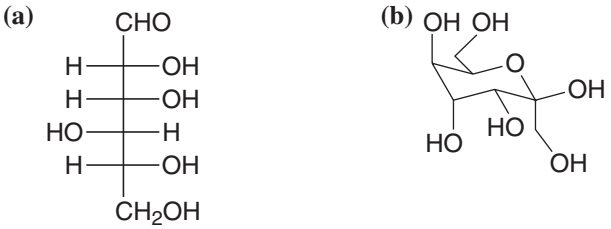
**These names come from the names of the oxygen heterocycles *pyran* and *furan* + *ose*:



A pyran **Furan**

Review Problem 22.3

Draw the β-pyranose form of (a) in its lowest energy chair conformation, and a Fischer projection for (b).

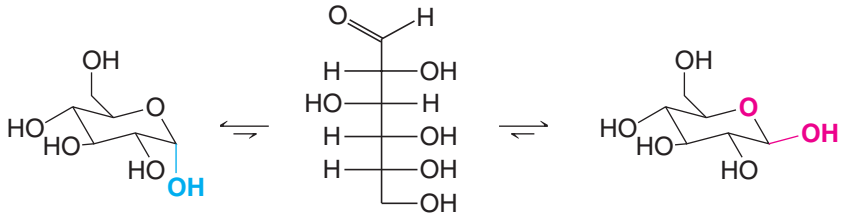


22.3 Mutarotation

Part of the evidence for the cyclic hemiacetal structure for D-(+)-glucose comes from experiments in which both α and β forms have been isolated. Ordinary D-(+)-glucose has a melting point of 146°C. However, when D-(+)-glucose is crystallized by evaporating an aqueous solution kept above 98°C, a second form of D-(+)-glucose with a melting point of 150°C can be obtained. When the optical rotations of these two forms are measured, they are found to be significantly different, but when an aqueous solution of either form is allowed to stand, its rotation changes. The specific rotation of one form decreases and the rotation of the other increases, *until both solutions show the same value*. A solution of ordinary D-(+)-glucose (mp 146°C) has an initial specific rotation of +112, but, ultimately, the specific rotation of this solution falls to +52.7. A solution of the second form of D-(+)-glucose (mp 150°C) has an initial specific rotation of +18.7, but, slowly, the specific rotation of this solution rises to +52.7.

- This change in specific rotation toward an equilibrium value is called **mutarotation**.

The explanation for this mutarotation lies in the existence of an equilibrium between the open-chain form of D-(+)-glucose and the α and β forms of the cyclic hemiacetals:

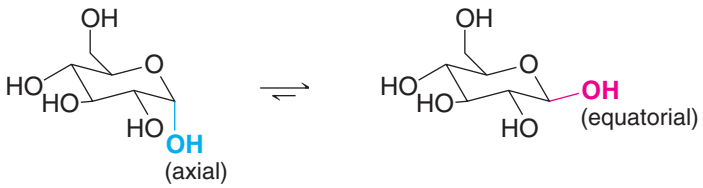


α-D-(+)-Glucopyranose (mp 146°C; $[\alpha]_D^{25}$ = +112) ⇌ **Open-chain form of D-(+)-glucose** ⇌ **β-D-(+)-Glucopyranose (mp 150°C; $[\alpha]_D^{25}$ = +18.7)**

X-Ray analysis has confirmed that ordinary D-(+)-glucose has the α configuration at the anomeric carbon atom and that the higher melting form has the β configuration.

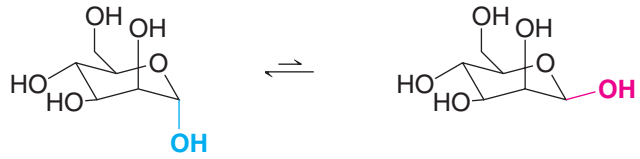
The concentration of open-chain D-(+)-glucose in solution at equilibrium is very small. Solutions of D-(+)-glucose give no observable UV or IR absorption band for a carbonyl group, and solutions of D-(+)-glucose give a negative test with Schiff's reagent—a special reagent that requires a relatively high concentration of a free aldehyde group (rather than a hemiacetal) in order to give a positive test.

Assuming that the concentration of the open-chain form is negligible, one can, by use of the specific rotations in the preceding figures, calculate the percentages of the α and β anomers present at equilibrium. These percentages, 36% α anomer and 64% β anomer, are in accord with a greater stability for β-D-(+)-glucopyranose. This preference is what we might expect on the basis of its having only equatorial groups:



α-D-(+)-Glucopyranose (36% at equilibrium) ⇌ **β-D-(+)-Glucopyranose (64% at equilibrium)**

The β anomer of a pyranose is not always the more stable, however. With D-mannose, the equilibrium favors the α anomer, and this result is called an *anomeric effect*:

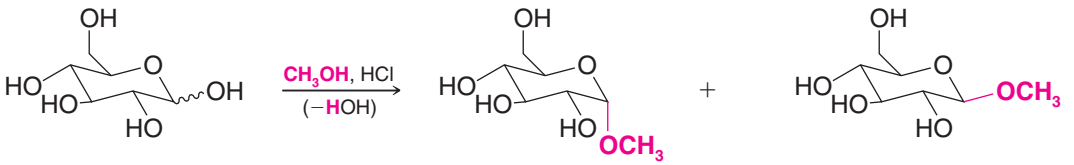


α-D-Mannopyranose (69% at equilibrium) ⇌ **β-D-Mannopyranose (31% at equilibrium)**

The anomeric effect is widely believed to be caused by hyperconjugation. An axially oriented orbital associated with nonbonding electrons of the ring oxygen can overlap with a σ* orbital of the axial exocyclic C—O hemiacetal bond. This effect is similar to that which causes the lowest energy conformation of ethane to be the anti conformation (Section 4.8). An anomeric effect will frequently cause an electronegative substituent, such as a hydroxyl or alkoxyl group, to prefer the axial orientation.

22.4 Glycoside Formation

When a small amount of gaseous hydrogen chloride is passed into a solution of D-(+)-glucose in methanol, a reaction takes place that results in the formation of anomeric methyl *acetals*:



D-(+)-Glucose → **Methyl α-D-glucopyranoside (mp 165°C; $[\alpha]_D^{25} = +158$)** + **Methyl β-D-glucopyranoside (mp 107°C; $[\alpha]_D^{25} = -33$)**

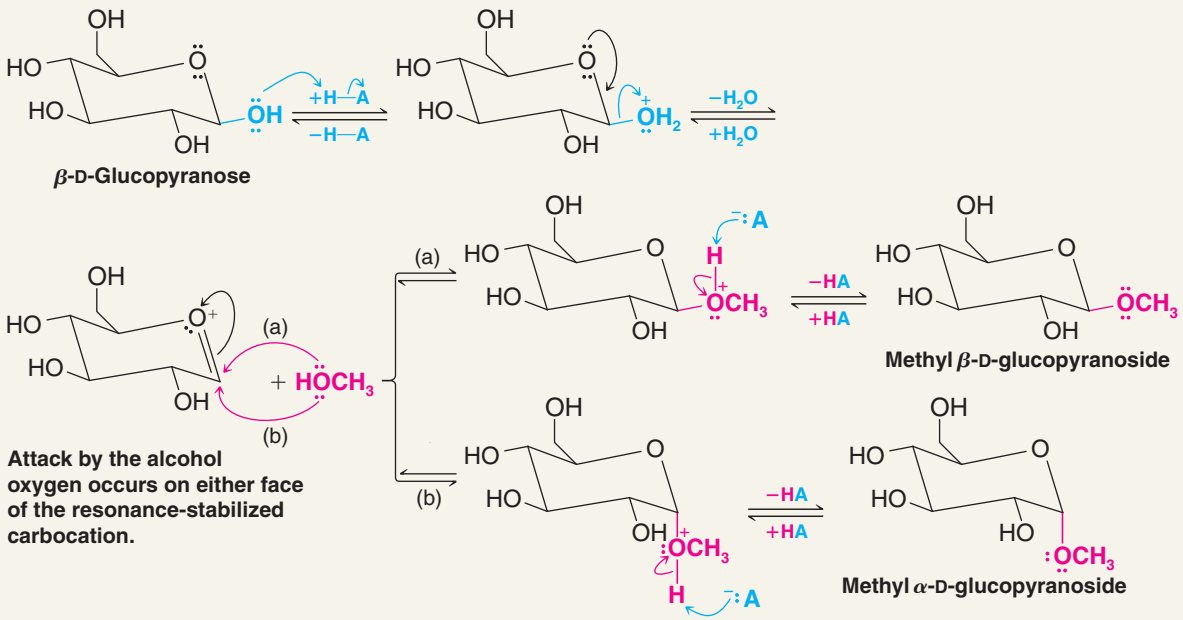
- Carbohydrate acetals are generally called **glycosides** (see the following mechanism), and an acetal of glucose is called a *glucoside*. (Acetals of mannose are *mannosides*, acetals of fructose are *fructosides*, and so on.)

The methyl D-glucosides have been shown to have six-membered rings (Section 22.2C) so they are properly named methyl α-D-glucopyranoside and methyl β-D-glucopyranoside.

The mechanism for the formation of the methyl glucosides (starting arbitrarily with β-D-glucopyranose) is as follows:

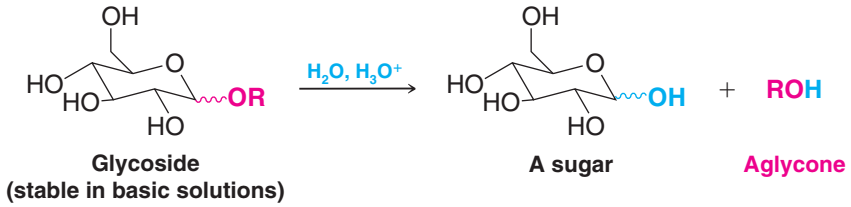
A MECHANISM FOR THE REACTION

Formation of a Glycoside



You should review the mechanism for acetal formation given in Section 16.7B and compare it with the steps given here. Notice, again, the important role played by the electron pair of the adjacent oxygen atom in stabilizing the carbocation that forms in the second step.

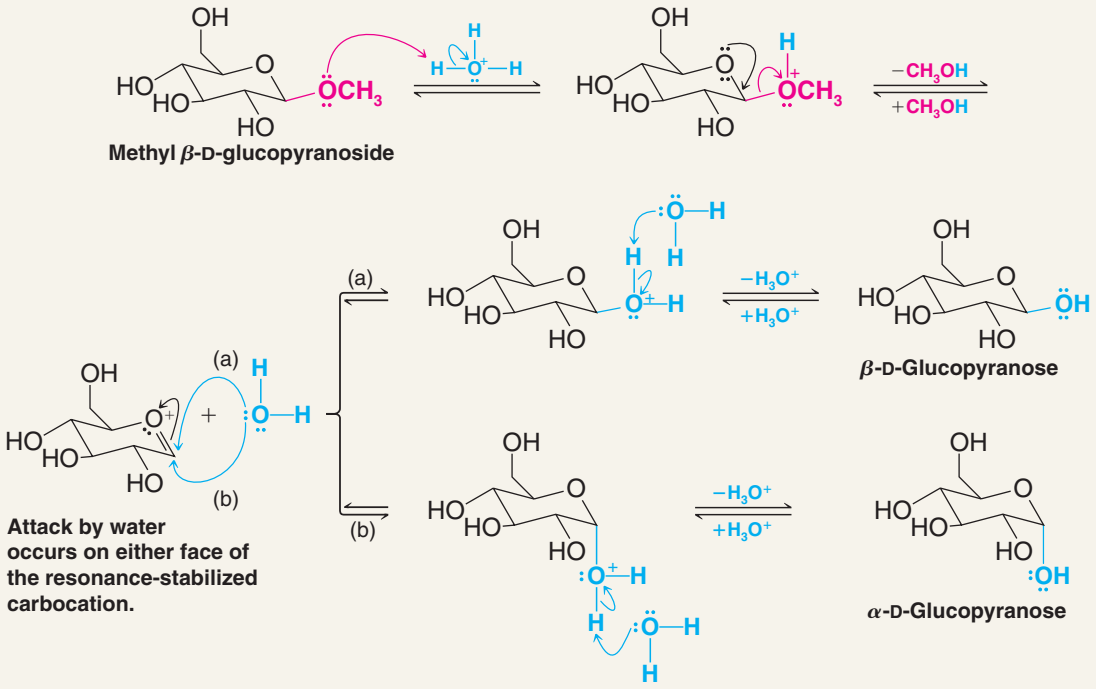
Glycosides are stable in basic solutions because they are acetals. In acidic solutions, however, glycosides undergo hydrolysis to produce a sugar and an alcohol (again, because they are acetals, Section 16.7B). The alcohol obtained by hydrolysis of a glycoside is known as an **aglycone**:



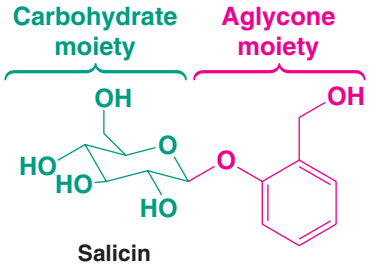
For example, when an aqueous solution of methyl β-D-glucopyranoside is made acidic, the glycoside undergoes hydrolysis to produce D-glucose as a mixture of the two pyranose forms (in equilibrium with a small amount of the open-chain form).

A MECHANISM FOR THE REACTION

Hydrolysis of a Glycoside



Glycosides may be as simple as the methyl glucosides that we have just studied or they may be considerably more complex. Many naturally occurring compounds are glycosides. An example is *salicin*, a compound found in the bark of willow trees:



As early as the time of the ancient Greeks, preparations made from willow bark were used in relieving pain. Eventually, chemists isolated salicin from other plant materials and were able to show that it was responsible for the analgesic effect of the willow bark preparations. Salicin can be converted to salicylic acid, which in turn can be converted into the most widely used modern analgesic, *aspirin* (Section 21.8).

Solved Problem 22.1

In neutral or basic solutions, glycosides do not show mutarotation. However, if the solutions are made acidic, glycosides show mutarotation. Explain.

ANSWER Because glycosides are acetals, they undergo hydrolysis in aqueous acid to form cyclic hemiacetals that then undergo mutarotation. Acetals are stable to base, and therefore in basic solution they do not show mutarotation.

(a) What products would be formed if salicin were treated with dilute aqueous HCl?

(b) Outline a mechanism for the reactions involved in their formation.

Review Problem 22.4

How would you convert D-glucose to a mixture of ethyl α-D-glucopyranoside and ethyl β-D-glucopyranoside? Show all steps in the mechanism for their formation.

Review Problem 22.5

22.5 Other Reactions of Monosaccharides

22.5A Enolization, Tautomerization, and Isomerization

Dissolving monosaccharides in aqueous base causes them to undergo a series of enolizations and keto–enol tautomerizations that lead to isomerizations. For example, if a solution of D-glucose containing calcium hydroxide is allowed to stand for several days, a number of products can be isolated, including D-fructose and D-mannose (Fig. 22.5). This type of reaction is called the **Lobry de Bruyn–Alberda van Ekenstein transformation** after the two Dutch chemists who discovered it in 1895.

When carrying out reactions with monosaccharides, it is usually important to prevent these isomerizations and thereby to preserve the stereochemistry at all of the chirality centers. One way to do this is to convert the monosaccharide to the methyl glycoside first. We can then safely carry out reactions in basic media because the aldehyde group has been converted to an acetal and acetals are stable in aqueous base. Preparation of the methyl glycoside serves to "protect" the monosaccharide from undesired reactions that could occur with the anomeric carbon in its hemiacetal form.

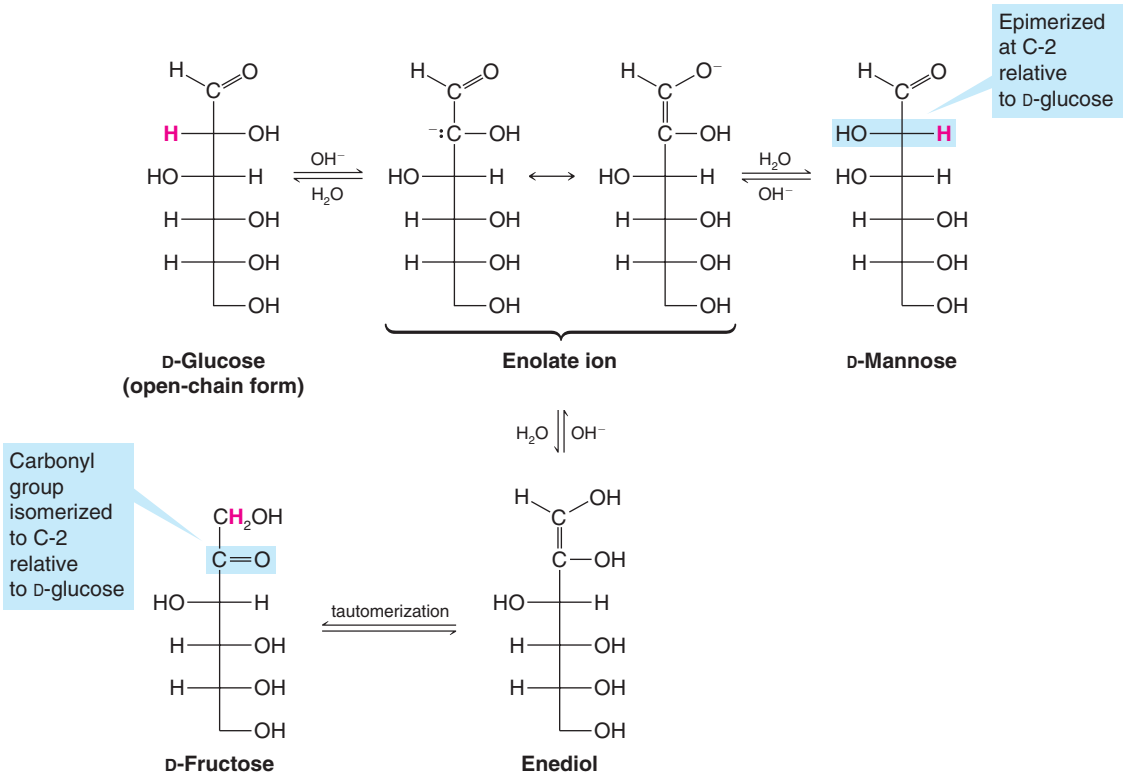


FIGURE 22.5 Monosaccharides undergo isomerizations via enolates and enediols when placed in aqueous base. Here we show how D-glucose isomerizes to D-mannose and to D-fructose.

22.5B Use of Protecting Groups in Carbohydrate Synthesis

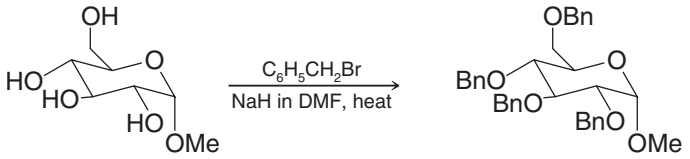
Protecting groups are functional groups introduced selectively to block the reactivity of certain sites in a molecule while desired transformations are carried on elsewhere. After the desired transformations are accomplished, the protecting groups are removed. Laboratory reactions involving carbohydrates often require the use of protecting groups due to the multiple sites of reactivity present in carbohydrates. As we have just seen, formation of a glycoside (an acetal) can be used to prevent undesired reactions that would involve the anomeric carbon in its hemiacetal form. Common protecting groups for the alcohol functional groups in carbohydrates include ethers, esters, and acetals.

22.5C Formation of Ethers

- Hydroxyl groups of sugars can be converted to ethers using a base and an alkyl halide by a version of the Williamson ether synthesis (Section 11.11B).

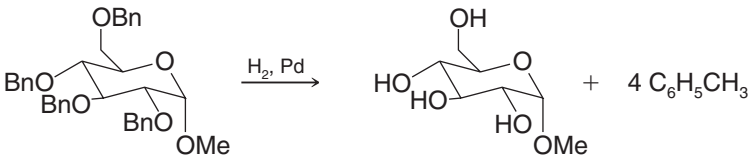
Benzyl ethers are commonly used to protect hydroxyl groups in sugars. Benzyl halides are easily introduced because they are highly reactive in S_N2 reactions. Sodium or potassium hydride is typically used as the base in an aprotic solvent such as DMF or DMSO. The benzyl groups can later be easily removed by hydrogenolysis using a palladium catalyst.

Benzyl Ether Formation

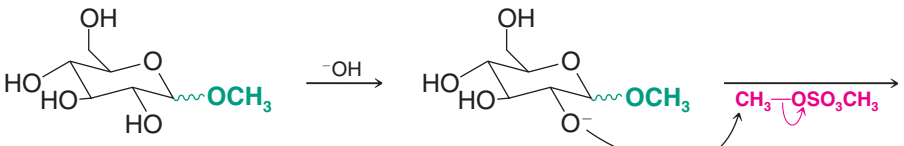


Bn = $C_6H_5CH_2$

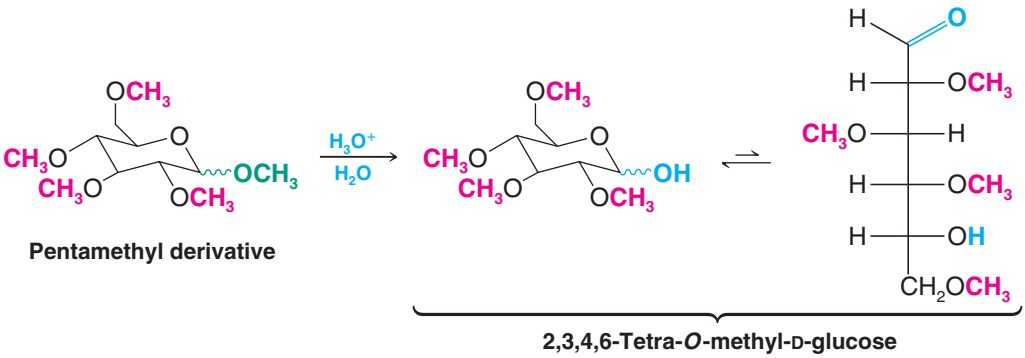
Benzyl Ether Cleavage



Methyl ethers can also be prepared. The pentamethyl derivative of glucopyranose, for example, can be synthesized by treating methyl glucoside with excess dimethyl sulfate in aqueous sodium hydroxide. Sodium hydroxide is a competent base in this case because the hydroxyl groups of monosaccharides are more acidic than those of ordinary alcohols due to the many electronegative atoms in the sugar, all of which exert electron-withdrawing inductive effects on nearby hydroxyl groups. In aqueous NaOH the hydroxyl groups are all converted to alkoxide ions, and each of these, in turn, reacts with dimethyl sulfate in an S_N2 reaction to yield a methyl ether. The process is called *exhaustive methylation*:

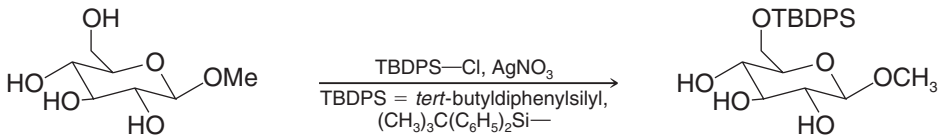


Although not often used as protecting groups for alcohols in carbohydrates, methyl ethers have been useful in the structure elucidation of sugars. For example, evidence for the pyranose form of glucose can be obtained by exhaustive methylation followed by aqueous hydrolysis of the acetal linkage. Because the C2, C3, C4, and C6 methoxy groups of the pentamethyl derivative are ethers, they are not affected by aqueous hydrolysis. (To cleave them requires heating with concentrated HBr or HI, Section 11.12.) The methoxyl group at C1, however, is part of an acetal linkage, and so it is labile under the conditions of aqueous hydrolysis. Hydrolysis of the pentamethyl derivative of glucose gives evidence that the C5 oxygen was the one involved in the cyclic hemiacetal form because in the open-chain form of the product (which is in equilibrium with the cyclic hemiacetal) it is the C5 oxygen that is not methylated:

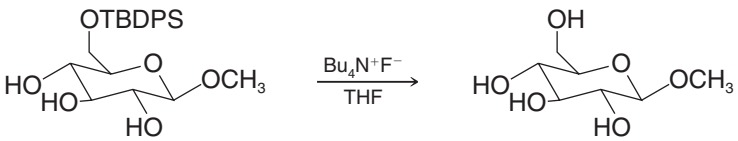


Silyl ethers, including *tert*-butyldimethylsilyl (TBS) ethers (Section 11.11E) and phenyl-substituted ethers, are also used as protecting groups in carbohydrate synthesis. *tert*-Butyldiphenylsilyl (TBDPS) ethers show excellent regioselectivity for primary hydroxyl groups in sugars, such as at C6 in a hexopyranose. (We shall see the use of some related silyl ether groups in Section 22.13D.)

Regioselective TBDPS Ether Formation

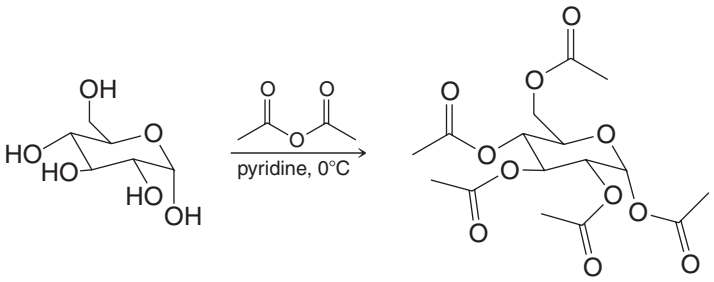


TBDPS Ether Cleavage



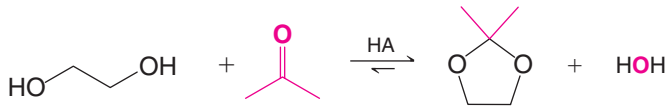
22.5D Conversion to Esters

Treating a monosaccharide with excess acetic anhydride and a weak base (such as pyridine or sodium acetate) converts all of the hydroxyl groups, including the anomeric hydroxyl, to ester groups. If the reaction is carried out at a low temperature (e.g., 0°C), the reaction occurs stereospecifically; the α anomer gives the α-acetate and the β anomer gives the β-acetate. Acetate esters are common protecting groups for carbohydrate hydroxyls.



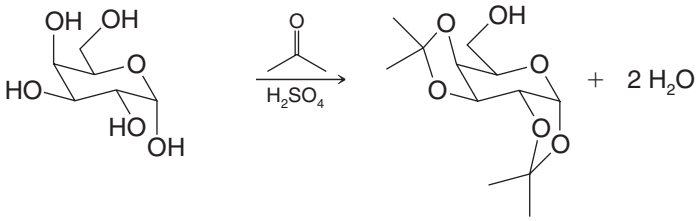
22.5E Conversion to Cyclic Acetals

In Section 16.7B we learned that aldehydes and ketones react with open-chain 1,2-diols to produce **cyclic acetals**:



1,2-Diol **Cyclic acetal**

If the 1,2-diol is attached to a ring, as in a monosaccharide, **formation of the cyclic acetals occurs only when the vicinal hydroxyl groups are cis to each other**. For example, α-D-galactopyranose reacts with acetone in the following way:



Cyclic acetals are commonly used to protect vicinal cis hydroxyl groups of a sugar while reactions are carried out on other parts of the molecule. When acetals such as these are formed from acetone, they are called **acetonides**.

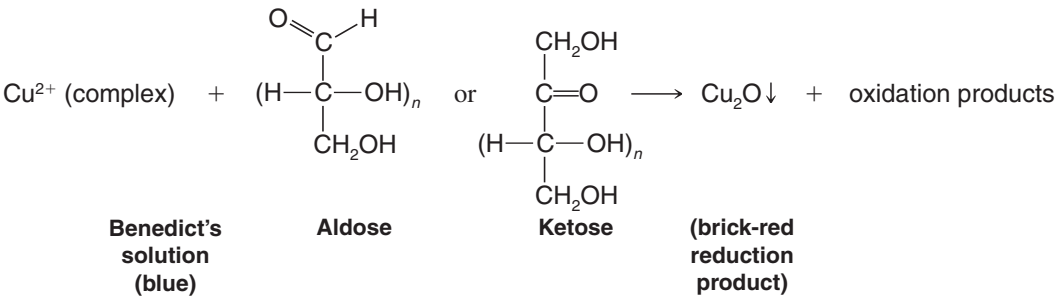
22.6 Oxidation Reactions of Monosaccharides

A number of oxidizing agents are used to identify functional groups of carbohydrates, in elucidating their structures, and for syntheses. The most important are (1) Benedict's or Tollens' reagents, (2) bromine water, (3) nitric acid, and (4) periodic acid. Each of these reagents produces a different and usually specific effect when it is allowed to react with a monosaccharide. We shall now examine what these effects are.

22.6A Benedict's or Tollens' Reagents: Reducing Sugars

Benedict's reagent (an alkaline solution containing a cupric citrate complex ion) and Tollens' solution $[Ag(NH_3)_2OH]$ oxidize and thus give positive tests with *aldoses and ketoses*. The tests are positive even though aldoses and ketoses exist primarily as cyclic hemiacetals.

We studied the use of Tollens' silver mirror test in Section 16.12B. Benedict's solution and the related Fehling's solution (which contains a cupric tartrate complex ion) give brick-red precipitates of Cu_2O when they oxidize an aldose. [In alkaline solution ketoses are converted to aldoses (Section 22.5A), which are then oxidized by the cupric complexes.] Since the solutions of cupric tartrates and citrates are blue, the appearance of a brick-red precipitate is a vivid and unmistakable indication of a positive test.

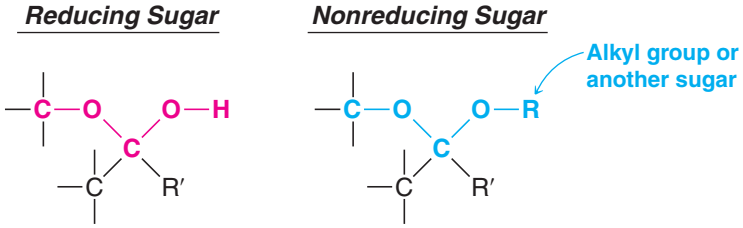


- Sugars that give positive tests with Tollens' or Benedict's solutions are known as **reducing sugars**, and all carbohydrates that contain a *hemiacetal group* give positive tests.

In aqueous solution the hemiacetal form of sugars exists in equilibrium with relatively small, but not insignificant, concentrations of noncyclic aldehydes or α-hydroxy ketones. It is the latter two that undergo the oxidation, perturbing the equilibrium to produce more aldehyde or α-hydroxy ketone, which then undergoes oxidation until one reactant is exhausted.

- Carbohydrates that contain only acetal groups do not give positive tests with Benedict's or Tollens' solutions, and they are called *nonreducing sugars*.

Acetals do not exist in equilibrium with aldehydes or α-hydroxy ketones in the basic aqueous media of the test reagents.



How might you distinguish between α-D-glucopyranose (i.e., D-glucose) and methyl α-D-glucopyranoside?

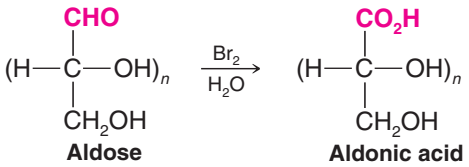
Review Problem 22.6

Although Benedict's and Tollens' reagents have some use as diagnostic tools [Benedict's solution can be used in quantitative determinations of reducing sugars (reported as glucose) in blood or urine], neither of these reagents is useful as a preparative reagent in carbohydrate oxidations. Oxidations with both reagents take place in alkaline solution, *and in alkaline solutions sugars undergo a complex series of reactions that lead to isomerizations* (Section 22.5A).

22.6B Bromine Water: The Synthesis of Aldonic Acids

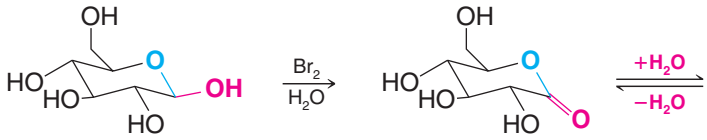
Monosaccharides do not undergo isomerization and fragmentation reactions in mildly acidic solution. Thus, a useful oxidizing reagent for preparative purposes is bromine in water (pH 6.0).

- Bromine water is a general reagent that selectively oxidizes the —CHO group to a —CO_2H group, thus converting an aldose to an **aldonic acid**:

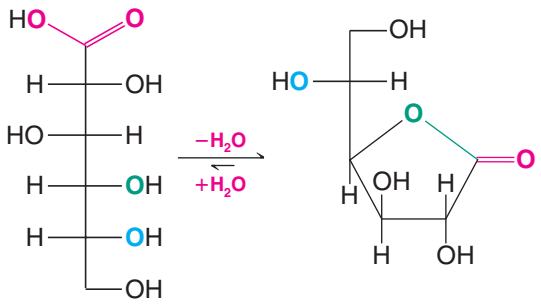


Aldose **Aldonic acid**

Experiments with aldopyranoses have shown that the actual course of the reaction is somewhat more complex than we have indicated. Bromine water specifically oxidizes the β anomer, and the initial product that forms is a δ-*aldonolactone*. This compound may then hydrolyze to an aldonic acid, and the aldonic acid may undergo a subsequent ring closure to form a γ-*aldonolactone:*



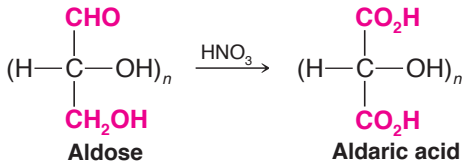
β-D-Glucopyranose **D-Glucono-δ-lactone**



D-Gluconic acid **D-Gluconic-γ-lactone**

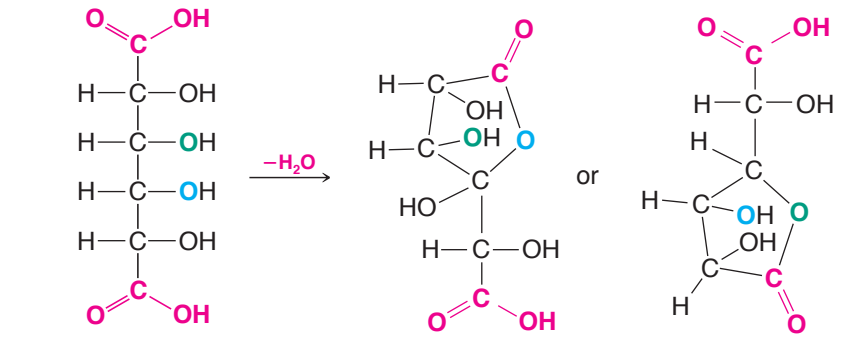
22.6C Nitric Acid Oxidation: Aldaric Acids

- Dilute nitric acid—a stronger oxidizing agent than bromine water—oxidizes both the —CHO group and the terminal —CH_2OH group of an aldose to —CO_2H groups, forming dicarboxylic acids are known as **aldaric acids**:



Aldose **Aldaric acid**

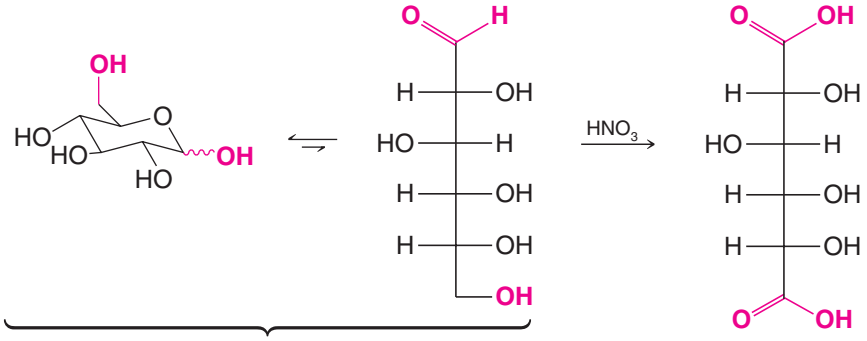
It is not known whether a lactone is an intermediate in the oxidation of an aldose to an aldaric acid; however, aldaric acids form γ- and δ-lactones readily:



Aldaric acid (from an aldohexose)

γ-Lactones of an aldaric acid

The aldaric acid obtained from D-glucose is called D-glucaric acid*:



D-Glucose

D-Glucaric acid

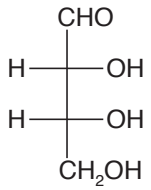
Review Problem 22.7

(a) Would you expect D-glucaric acid to be optically active?

(b) Write the open-chain structure for the aldaric acid (mannaric acid) that would be obtained by nitric acid oxidation of D-mannose.

(c) Would you expect mannaric acid to be optically active?

(d) What aldaric acid would you expect to obtain from D-erythrose?



D-Erythrose

(e) Would the aldaric acid in (d) show optical activity?

(f) D-Threose, a diastereomer of D-erythrose, yields an optically active aldaric acid when it is subjected to nitric acid oxidation. Write Fischer projection formulas for D-threose and its nitric acid oxidation product.

(g) What are the names of the aldaric acids obtained from D-erythrose and D-threose?

*Older terms for an aldaric acid are a *glycaric* acid or a *saccharic* acid.

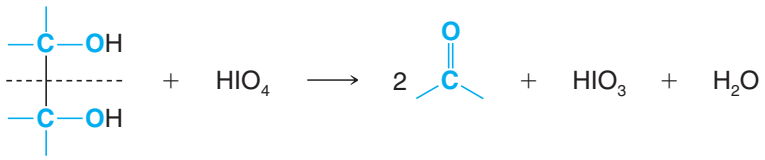
Review Problem 22.8

D-Glucaric acid undergoes lactonization to yield two different γ-lactones. What are their structures?

22.6D Periodate Oxidations: Oxidative Cleavage of Polyhydroxy Compounds

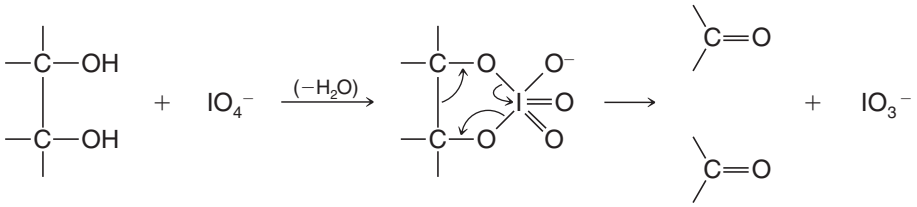
- Compounds that have hydroxyl groups on adjacent atoms undergo oxidative cleavage when they are treated with aqueous periodic acid (HIO_4). The reaction breaks carbon–carbon bonds and produces carbonyl compounds (aldehydes, ketones, or acids).

The stoichiometry of oxidative cleavage by periodic acid is



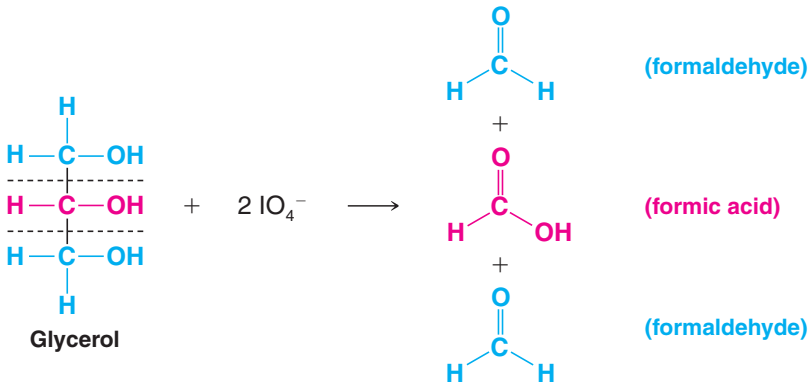
Since the reaction usually takes place in quantitative yield, valuable information can often be gained by measuring the number of molar equivalents of periodic acid consumed in the reaction as well as by identifying the carbonyl products.*

Periodate oxidations are thought to take place through a cyclic intermediate:



Before we discuss the use of periodic acid in carbohydrate chemistry, we should illustrate the course of the reaction with several simple examples. Notice in these periodate oxidations that *for every C—C bond broken, a C—O bond is formed at each carbon*.

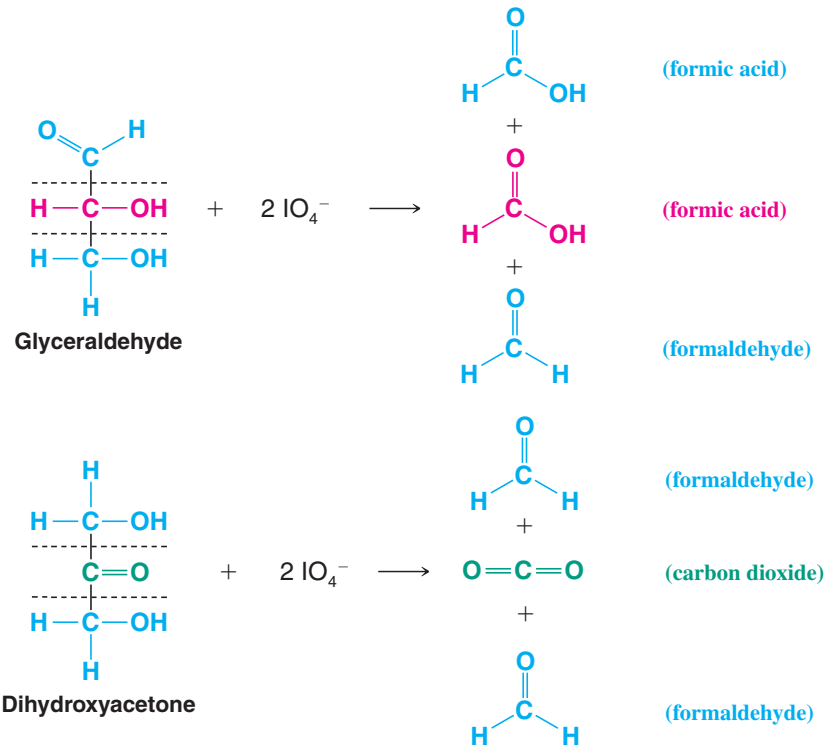
1. When three or more —CHOH groups are contiguous, the internal ones are obtained as *formic acid*. Periodate oxidation of glycerol, for example, gives two molar equivalents of formaldehyde and one molar equivalent of formic acid:



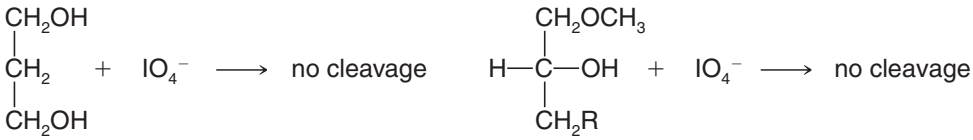
2. Oxidative cleavage also takes place when an —OH group is adjacent to the carbonyl group of an aldehyde or ketone (but not that of an acid or an ester). Glyceraldehyde yields two molar equivalents of formic acid and one molar equivalent of formalde-

*The reagent lead tetraacetate, $Pb(O_2CCH_3)_4$, brings about cleavage reactions similar to those of periodic acid. The two reagents are complementary; periodic acid works well in aqueous solutions and lead tetraacetate gives good results in organic solvents but is more toxic.

hyde, while dihydroxyacetone gives two molar equivalents of formaldehyde and one molar equivalent of carbon dioxide:



3. Periodic acid does not cleave compounds in which the hydroxyl groups are separated by an intervening —CH_2— group, nor those in which a hydroxyl group is adjacent to an ether or acetal function:



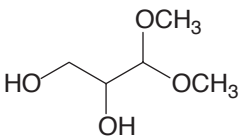
Review Problem 22.9

What products would you expect to be formed when each of the following compounds is treated with an appropriate amount of periodic acid? How many molar equivalents of HIO_4 would be consumed in each case?

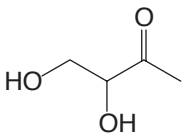
(a) 2,3-Butanediol

(b) 1,2,3-Butanetriol

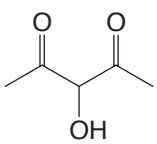
(c)



(d)

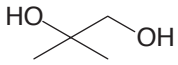


(e)



(f) *cis*-1,2-Cyclopentanediol

(g)



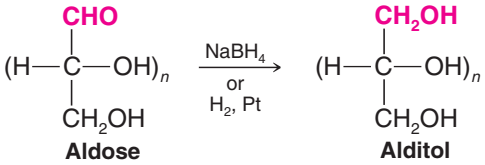
(h) D-Erythrose

Review Problem 22.10

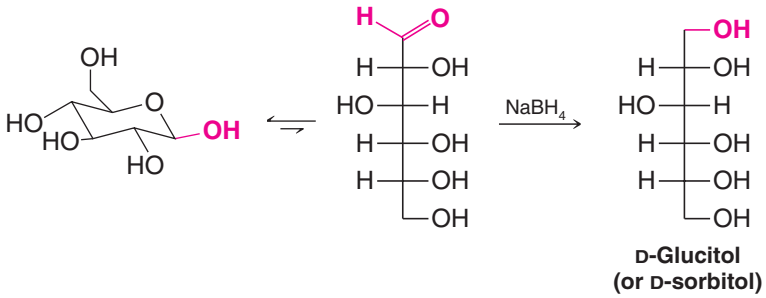
Show how periodic acid could be used to distinguish between an aldohexose and a ketohexose. What products would you obtain from each, and how many molar equivalents of HIO_4 would be consumed?

22.7 Reduction of Monosaccharides: Alditols

- Aldoses (and ketoses) can be reduced with sodium borohydride to compounds called **alditols**:



Reduction of D-glucose, for example, yields D-glucitol:

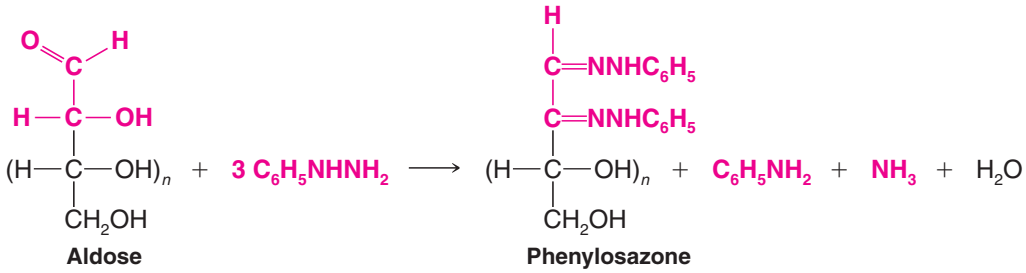


Review Problem 22.11

(a) Would you expect D-glucitol to be optically active? **(b)** Write Fischer projection formulas for all of the D-aldohexoses that would yield *optically inactive alditols*.

22.8 Reactions of Monosaccharides with Phenylhydrazine: Osazones

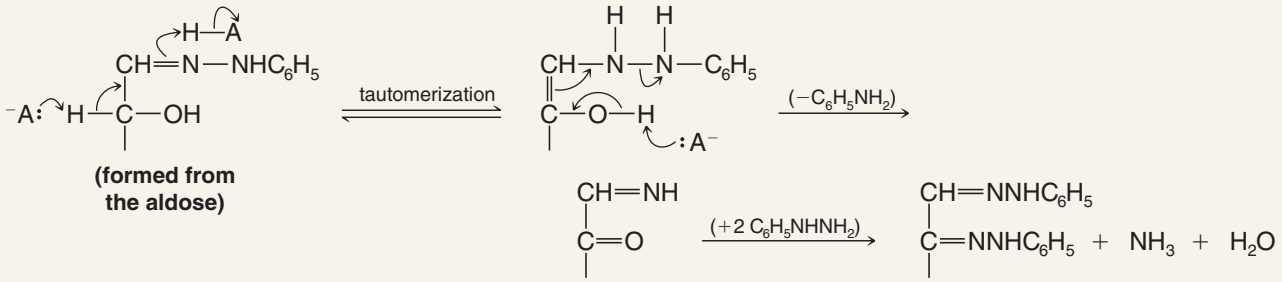
The aldehyde group of an aldose reacts with such carbonyl reagents as hydroxylamine and phenylhydrazine (Section 16.8B). With hydroxylamine, the product is the expected oxime. With enough phenylhydrazine, however, three molar equivalents of phenylhydrazine are consumed and a second phenylhydrazone group is introduced at C2. The product is called a *phenylosazone*. Phenylosazones crystallize readily (unlike sugars) and are useful derivatives for identifying sugars.



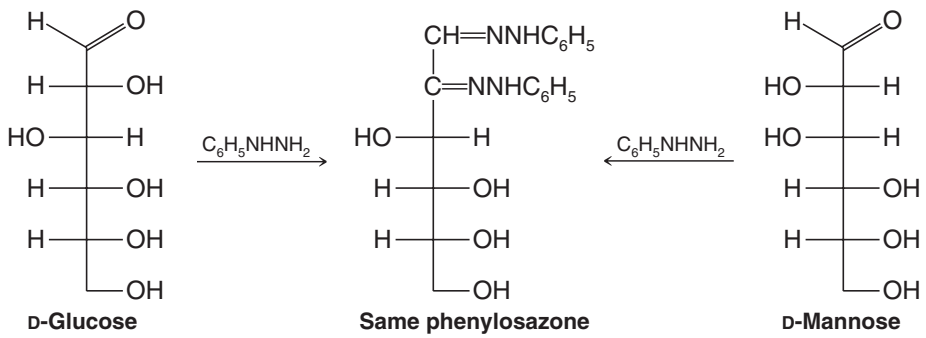
The mechanism for osazone formation probably depends on a series of reactions in which $\diagup C{=}N \diagdown$ behaves very much like $\diagup C{=}O \diagdown$ in giving a nitrogen version of an enol.

A MECHANISM FOR THE REACTION

Phenylosazone Formation



Osazone formation results in a loss of the chirality center at C2 but does not affect other chirality centers; D-glucose and D-mannose, for example, yield the same phenylosazone:



This experiment, first done by Emil Fischer, established that D-glucose and D-mannose have the same configurations about C3, C4, and C5. Diastereomeric aldoses that differ in configuration at only one carbon (such as D-glucose and D-mannose) are called epimers. In general, any pair of diastereomers that differ in configuration at only a single tetrahedral chirality center can be called **epimers**.

Review Problem 22.12

Although D-fructose is not an epimer of D-glucose or D-mannose (D-fructose is a ketohexose), all three yield the same phenylosazone. **(a)** Using Fischer projection formulas, write an equation for the reaction of fructose with phenylhydrazine. **(b)** What information about the stereochemistry of D-fructose does this experiment yield?

22.9 Synthesis and Degradation of Monosaccharides

22.9A Kiliani–Fischer Synthesis

In 1885, Heinrich Kiliani (Freiburg, Germany) discovered that an aldose can be converted to the epimeric aldonic acids having one additional carbon through the addition of hydrogen cyanide and subsequent hydrolysis of the epimeric cyanohydrins. Fischer later extended this method by showing that aldonolactones obtained from the aldonic acids can be reduced to aldoses. Today, this method for lengthening the carbon chain of an aldose is called the Kiliani–Fischer synthesis.

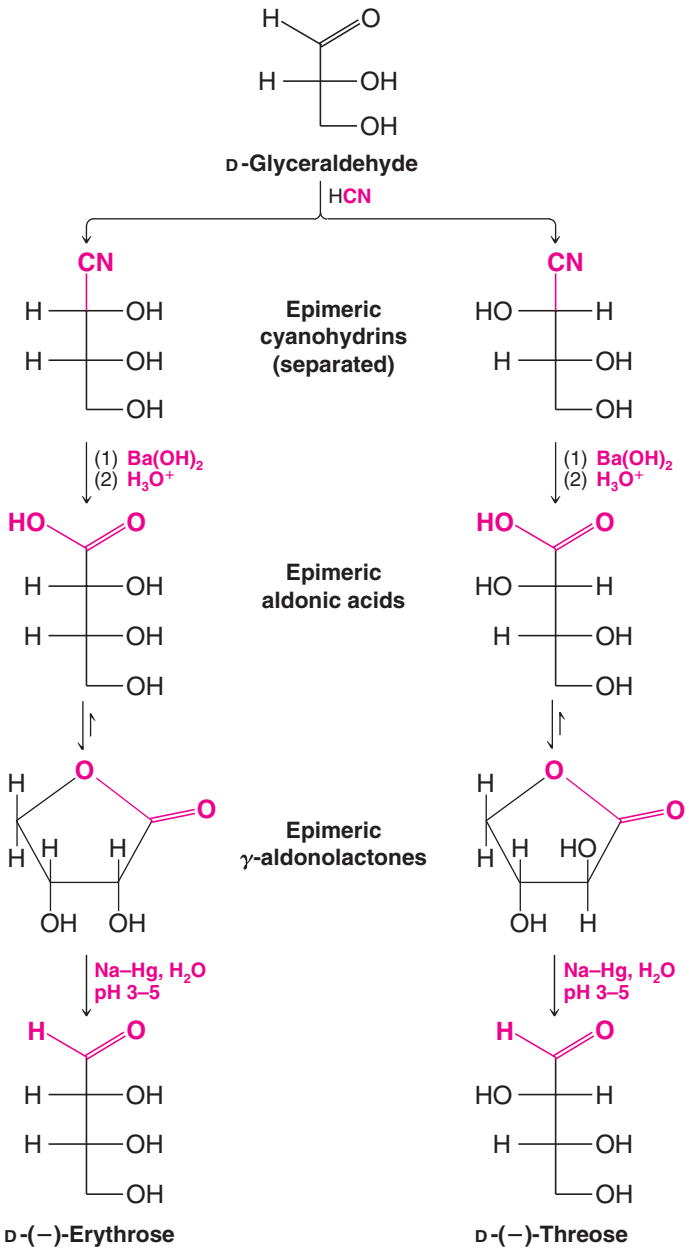
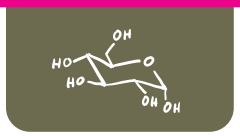


Figure 22.6 A Kiliani–Fischer synthesis of D-(−)-erythrose and D-(−)-threose from D-glyceraldehyde.

We can illustrate the Kiliani–Fischer synthesis with the synthesis of D-threose and D-erythrose (aldotetroses) from D-glyceraldehyde (an aldotriose) in Fig. 22.6.

Addition of hydrogen cyanide to glyceraldehyde produces two epimeric cyanohydrins because the reaction creates a new chirality center. The cyanohydrins can be separated easily (since they are diastereomers), and each can be converted to an aldose through hydrolysis, acidification, lactonization, and reduction with Na–Hg at pH 3–5. One cyanohydrin ultimately yields D-(−)-erythrose and the other yields D-(−)-threose.

We can be sure that the aldotetroses that we obtain from this Kiliani–Fischer synthesis are both D sugars because the starting compound is D-glyceraldehyde and its chirality center is unaffected by the synthesis. On the basis of the Kiliani–Fischer synthesis, we cannot know just which aldotetrose has both —OH groups on the right and which has the top —OH on the left in the Fischer projection. However, if we oxidize both aldotetroses to aldaric acids, one [D-(−)-erythrose] will yield an *optically inactive* (meso) product while the other [D-(−)-threose] will yield a product that is *optically active* (see Review Problem 22.7).



Review Problem 22.13

(a) What are the structures of L-(+)-threose and L-(+)-erythrose? **(b)** What aldotriose would you use to prepare them in a Kiliani–Fischer synthesis?

Review Problem 22.14

(a) Outline a Kiliani–Fischer synthesis of epimeric aldopentoses starting with D-(−)-erythrose (use Fischer projections). **(b)** The two epimeric aldopentoses that one obtains are D-(−)-arabinose and D-(−)-ribose. Nitric acid oxidation of D-(−)-ribose yields an optically inactive aldaric acid, whereas similar oxidation of D-(−)-arabinose yields an optically active product. On the basis of this information alone, which Fischer projection represents D-(−)-arabinose and which represents D-(−)-ribose?

Review Problem 22.15

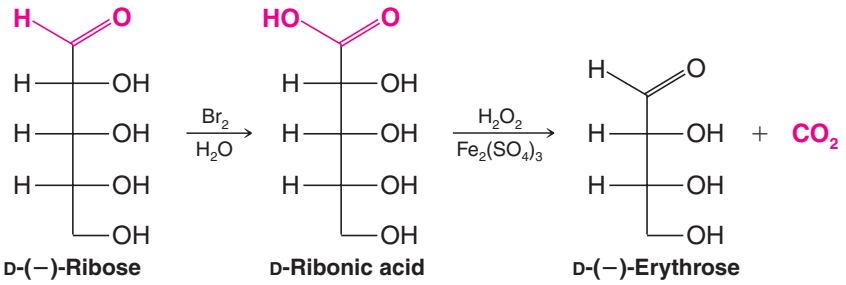
Subjecting D-(−)-threose to a Kiliani–Fischer synthesis yields two other epimeric aldopentoses, D-(+)-xylose and D-(−)-lyxose. D-(+)-Xylose can be oxidized (with nitric acid) to an optically inactive aldaric acid, while similar oxidation of D-(−)-lyxose gives an optically active product. What are the structures of D-(+)-xylose and D-(−)-lyxose?

Review Problem 22.16

There are eight aldopentoses. In Review Problems 22.14 and 22.15 you have arrived at the structures of four. What are the names and structures of the four that remain?

22.9B The Ruff Degradation

Just as the Kiliani–Fischer synthesis can be used to lengthen the chain of an aldose by one carbon atom, the Ruff degradation* can be used to shorten the chain by a similar unit. The Ruff degradation involves (1) oxidation of the aldose to an aldonic acid using bromine water and (2) oxidative decarboxylation of the aldonic acid to the next lower aldose using hydrogen peroxide and ferric sulfate. D-(−)-Ribose, for example, can be degraded to D-(−)-erythrose:



D-(−)-Ribose D-Ribonic acid D-(−)-Erythrose

Review Problem 22.17

The aldohexose D-(+)-galactose can be obtained by hydrolysis of *lactose*, a disaccharide found in milk. When D-(+)-galactose is treated with nitric acid, it yields an optically inactive aldaric acid. When D-(+)-galactose is subjected to Ruff degradation, it yields D-(−)-lyxose (see Review Problem 22.15). Using only these data, write the Fischer projection formula for D-(+)-galactose.

22.10 The D Family of Aldoses

The Ruff degradation and the Kiliani–Fischer synthesis allow us to place all of the aldoses into families or "family trees" based on their relation to D- or L-glyceraldehyde. Such a tree is constructed in Fig. 22.7 and includes the structures of the D-aldohexoses, **1–8**.

- Most, but not all, of the naturally occurring aldoses belong to the D family, with D-(+)-glucose being by far the most common.

*Developed by Otto Ruff, 1871–1939, a German chemist.

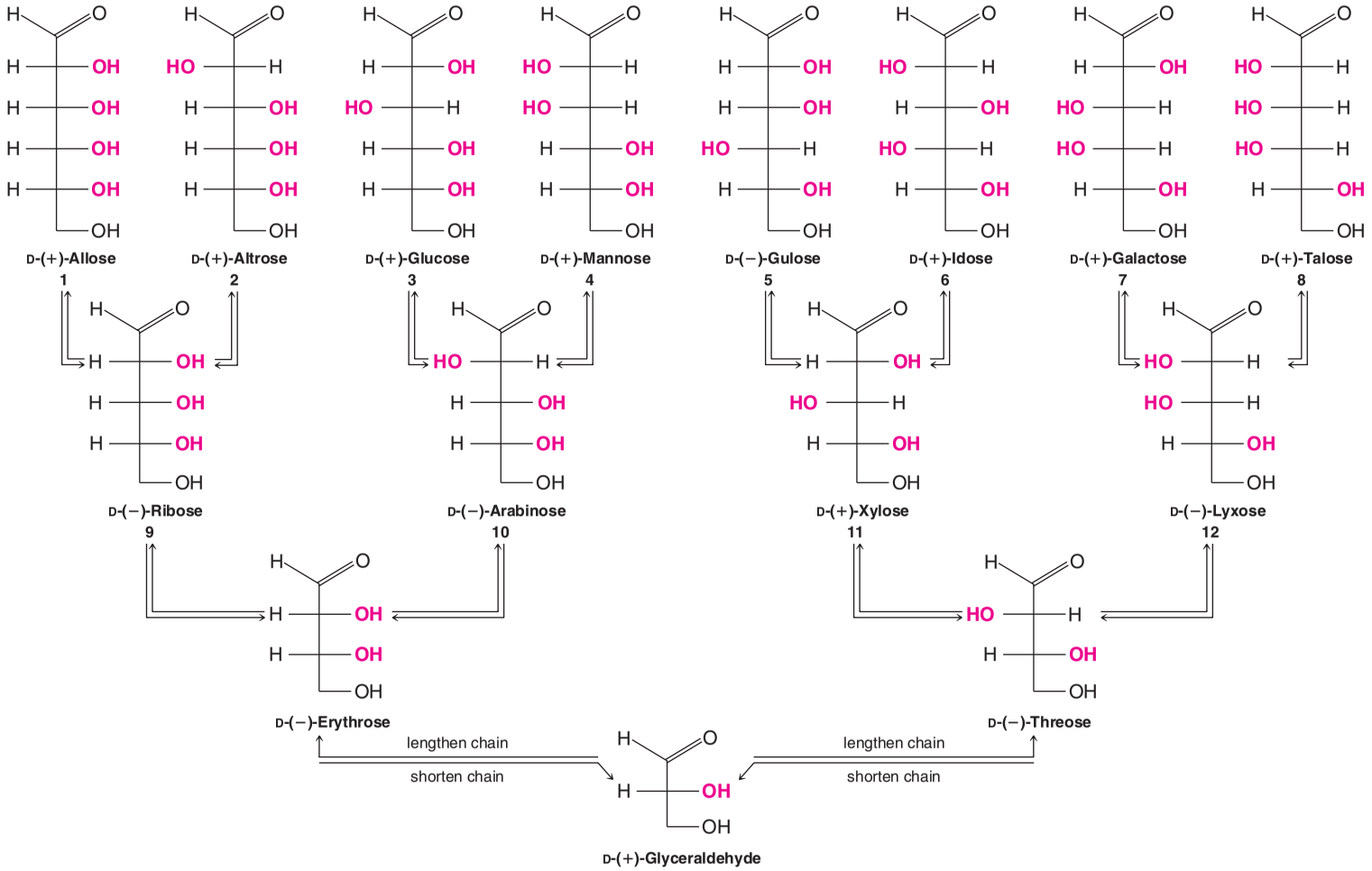


Figure 22.7 The D family of aldohexoses.* (From *Organic Chemistry* by Fieser, L. F., and Fieser, M. Copyright © 1956 by International Thompson.)

*A useful mnemonic for the D-aldohexoses: All altruists gladly make gum in gallon tanks. Write the names in a line and above each write CH_2OH. Then, for C5 write OH to the right all the way across. For C4 write OH to the right four times, then four to the left; for C3, write OH twice to the right, twice to the left, and repeat; for C2, alternate OH and H to the right. (From Fieser, L. F., and Fieser, M., *Organic Chemistry*, Reinhold: New York, 1956; p 359.)

D-(+)-Galactose can be obtained from milk sugar (lactose), but L-(−)-galactose occurs in a polysaccharide obtained from the vineyard snail, *Helix pomatia*. L-(+)-Arabinose is found widely, but D-(−)-arabinose is scarce, being found only in certain bacteria and sponges. Threose, lyxose, gulose, and allose do not occur naturally, but one or both forms (D or L) of each have been synthesized.

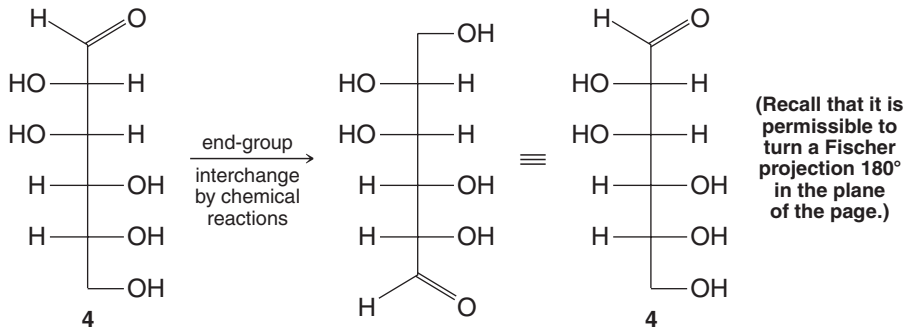
22.11 *Fischer's Proof of the Configuration of D-(+)-Glucose*

Emil Fischer began his work on the stereochemistry of (+)-glucose in 1888, only 12 years after van't Hoff and Le Bel had made their proposal concerning the tetrahedral structure of carbon. Only a small body of data was available to Fischer at the beginning: Only a few monosaccharides were known, including (+)-glucose, (+)-arabinose, and (+)-mannose. [(+)-Mannose had just been synthesized by Fischer.] The sugars (+)-glucose and (+)-mannose were known to be aldohexoses; (+)-arabinose was known to be an aldopentose.

Since an aldohexose has four chirality centers, 2^4 (or 16) stereoisomers are possible—*one of which is* (+)-*glucose*. Fischer arbitrarily decided to limit his attention to the eight structures with the D configuration given in Fig. 22.7 (structures **1–8**). Fischer realized that he would be unable to differentiate between enantiomeric configurations because methods for determining the absolute configuration of organic compounds had not been developed. It was not until 1951, when Bijvoet (Section 5.15A) determined the absolute configuration of L-(+)-tartaric acid [and, hence, D-(+)-glyceraldehyde], that Fischer's arbitrary assignment of (+)-glucose to the family we call the D family was known to be correct.

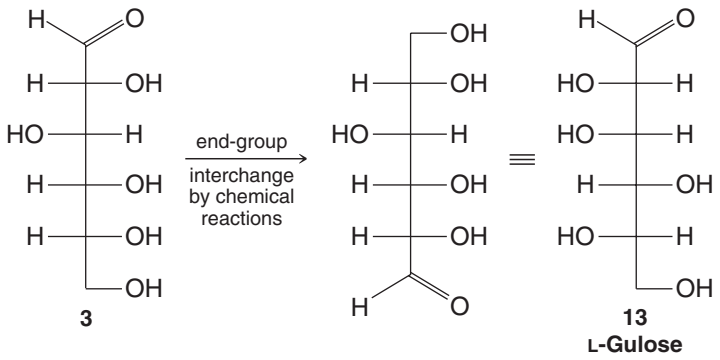
Fischer's assignment of structure **3** to (+)-glucose was based on the following reasoning:

1. Nitric acid oxidation of (+)-glucose gives an optically active aldaric acid. This eliminates structures **1** and **7** from consideration because both compounds would yield *meso*-aldaric acids.
2. *Degradation of* (+)-*glucose gives* (−)-*arabinose, and nitric acid oxidation of* (−)-*arabinose gives an optically active aldaric acid*. This means that (−)-arabinose cannot have configuration **9** or **11** and must have either structure **10** or **12**. It also establishes that (+)-glucose cannot have configuration **2, 5**, or **6**. This leaves structures **3, 4**, and **8** as possibilities for (+)-glucose.
3. Kiliani–Fischer synthesis beginning with (−)-arabinose gives (+)-glucose and (+)-mannose; nitric acid oxidation of (+)-mannose gives an optically active aldaric acid. This, together with the fact that (+)-glucose yields a different but also optically active aldaric acid, establishes **10** as the structure of (−)-arabinose and eliminates **8** as a possible structure for (+)-glucose. Had (−)-arabinose been represented by structure **12**, a Kiliani–Fischer synthesis would have given the two aldohexoses, **7** and **8**, one of which **(7)** would yield an optically inactive aldaric acid on nitric acid oxidation.
4. Two structures now remain, **3** and **4**; one structure represents (+)-glucose and one represents (+)-mannose. Fischer realized that (+)-glucose and (+)-mannose were epimeric (at C2), but a decision as to which compound was represented by which structure was most difficult.
5. Fischer had already developed a method for effectively *interchanging the two end groups* (aldehyde and primary alcohol) *of an aldose chain*. And, with brilliant logic, Fischer realized that if (+)-glucose had structure **4**, an interchange of end groups *would yield the same aldohexose*:



4 **4**

On the other hand, if (+)-glucose has structure **3**, *an end-group interchange will yield a different aldohexose*, **13**:

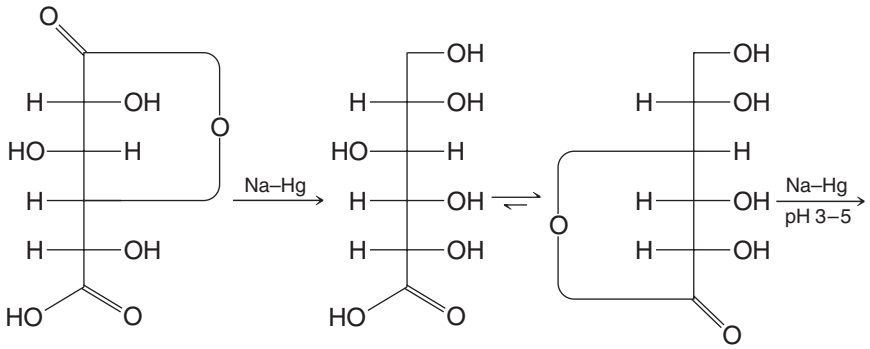


3 **13**
L-Gulose

This new aldohexose, if it were formed, would be an L sugar and it would be the mirror reflection of D-gulose. Thus its name would be L-gulose.

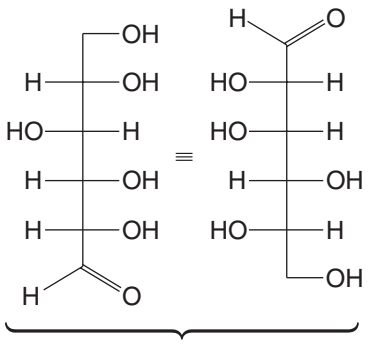
Fischer carried out the end-group interchange starting with (+)-glucose and *the product was the new aldohexose* **13**. This outcome proved that (+)-glucose has structure **3**. It also established **4** as the structure for (+)-mannose, and it proved the structure of L-(+)-gulose as **13**.

The procedure Fischer used for interchanging the ends of the (+)-glucose chain began with one of the γ-lactones of D-glucaric acid (see Review Problem 22.8) and was carried out as follows:



A γ-lactone of D-glucaric acid **L-Gulonic acid** **A γ-aldonolactone**

(continued from previous page)



L-(+)-Gulose
13

Helpful Hint

See WileyPLUS for "The Chemistry of... Stereoselective Synthesis of all the L-Aldohexoses."

Notice in this synthesis that the second reduction with Na–Hg is carried out at pH 3–5. Under these conditions, reduction of the lactone yields an aldehyde and not a primary alcohol.

Review Problem 22.18

Fischer actually had to subject both γ-lactones of D-glucaric acid (Review Problem 22.8) to the procedure just outlined. What product does the other γ-lactone yield?

22.12 Disaccharides

22.12A Sucrose

Ordinary table sugar is a disaccharide called *sucrose*. Sucrose, the most widely occurring disaccharide, is found in all photosynthetic plants and is obtained commercially from sugarcane or sugar beets. Sucrose has the structure shown in Fig. 22.8.

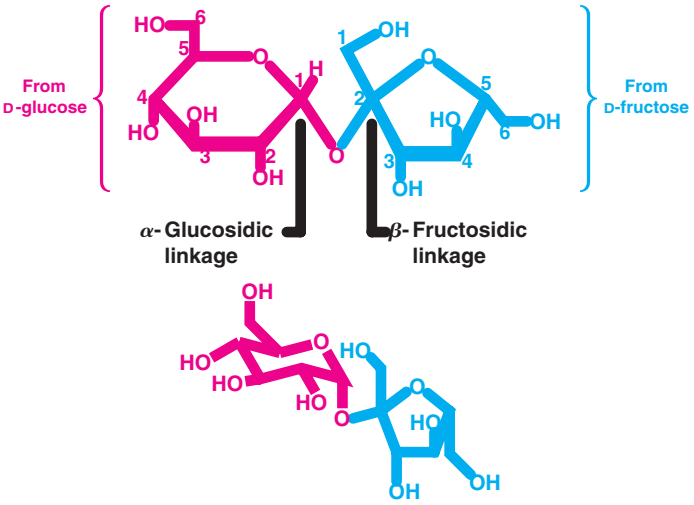
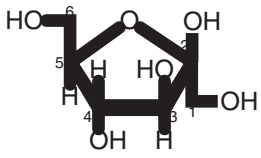


Figure 22.8 Two representations of the formula for (+)-sucrose (α-D-glucopyranosyl β-D-fructofuranoside).

The structure of sucrose is based on the following evidence:

1. Sucrose has the molecular formula $C_{12}H_{22}O_{11}$.
2. Acid-catalyzed hydrolysis of 1 mol of sucrose yields 1 mol of D-glucose and 1 mol of D-fructose.



Fructose
(as a β-furanose)

3. Sucrose is a nonreducing sugar; it gives negative tests with Benedict's and Tollens' solutions. Sucrose does not form an osazone and does not undergo mutarotation. These facts mean that neither the glucose nor the fructose portion of sucrose has a hemiacetal group. Thus, the two hexoses must have a glycosidic linkage that involves C1 of glucose and C2 of fructose, for only in this way will both carbonyl groups be present as full acetals (i.e., as glycosides).
4. The stereochemistry of the glycosidic linkages can be inferred from experiments done with enzymes. Sucrose is hydrolyzed by an *α-glucosidase* obtained from yeast but not by β-glucosidase enzymes. This hydrolysis indicates *an α configuration at the glucoside portion*. Sucrose is also hydrolyzed by *sucrase*, an enzyme known to hydrolyze β-fructofuranosides but not α-fructofuranosides. This hydrolysis indicates a *β configuration at the fructoside portion*.
5. Methylation of sucrose gives an octamethyl derivative that, on hydrolysis, gives 2,3,4,6-tetra-*O*-methyl-D-glucose and 1,3,4,6-tetra-*O*-methyl-D-fructose. The identities of these two products demonstrate that the glucose portion is a *pyranoside* and that the fructose portion is a *furanoside*.

The structure of sucrose has been confirmed by X-ray analysis and by an unambiguous synthesis.

22.12B Maltose

When starch (Section 22.13A) is hydrolyzed by the enzyme *diastase*, one product is a disaccharide known as *maltose* (Fig. 22.9). The structure of maltose was deduced based on the following evidence:

1. When 1 mol of maltose is subjected to acid-catalyzed hydrolysis, it yields 2 mol of D-(+)-glucose.
2. Unlike sucrose, *maltose is a reducing sugar*; it gives positive tests with Fehling's, Benedict's, and Tollens' solutions. Maltose also reacts with phenylhydrazine to form a monophenylosazone (i.e., it incorporates two molecules of phenylhydrazine).
3. Maltose exists in two anomeric forms: α-(+)-maltose, $[\alpha]_D^{25} = +168$, and β-(+)-maltose, $[\alpha]_D^{25} = +112$. The maltose anomers undergo mutarotation to yield an equilibrium mixture, $[\alpha]_D^{25} = +136$.

Facts 2 and 3 demonstrate that one of the glucose residues of maltose is present in a hemiacetal form; the other, therefore, must be present as a glucoside. The configuration of this glucosidic linkage can be inferred as α, because maltose is hydrolyzed by α-glucosidase enzymes and not by β-glucosidase enzymes.

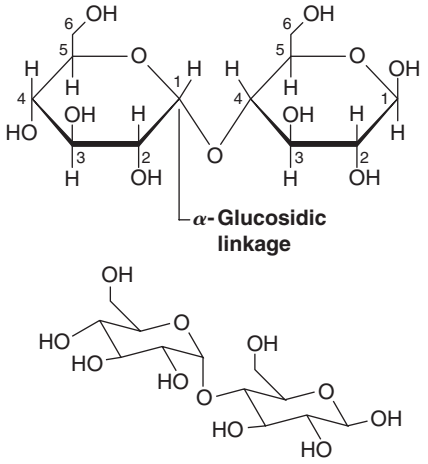


Figure 22.9 Two representations of the structure of the β anomer of (+)-maltose, 4-*O*-(α-D-glucopyranosyl)-β-D-glucopyranose.

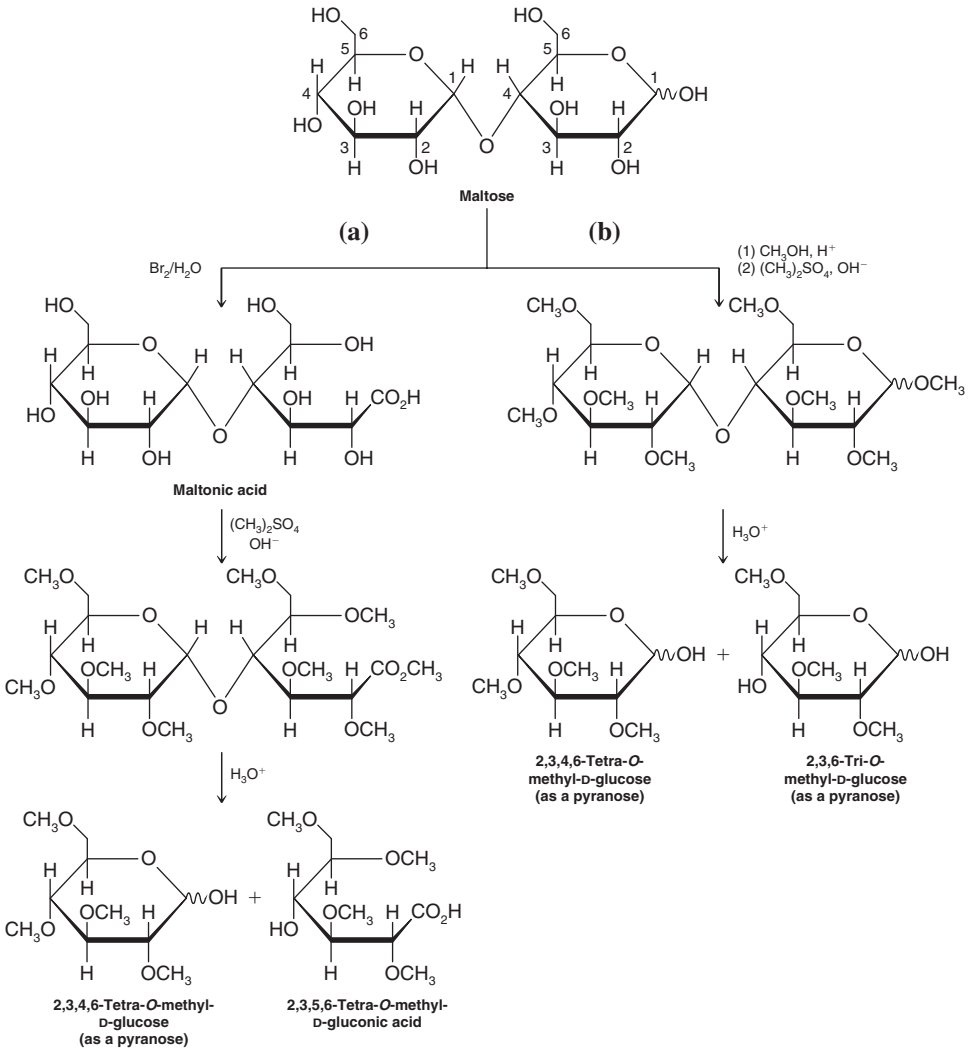


Figure 22.10 *(a)* Oxidation of maltose to maltonic acid followed by methylation and hydrolysis. *(b)* Methylation and subsequent hydrolysis of maltose itself.

4. Maltose reacts with bromine water to form a monocarboxylic acid, maltonic acid (Fig. 22.10*a*). This fact, too, is consistent with the presence of only one hemiacetal group.
5. Methylation of maltonic acid followed by hydrolysis gives 2,3,4,6-tetra-*O*-methyl-D-glucose and 2,3,5,6-tetra-*O*-methyl-D-gluconic acid. That the first product has a free —OH at C5 indicates that the nonreducing glucose portion is present as a pyranoside; that the second product, 2,3,5,6-tetra-*O*-methyl-D-gluconic acid, has a free —OH at C4 indicates that this position was involved in a glycosidic linkage with the nonreducing glucose. Only the size of the reducing glucose ring needs to be determined.
6. Methylation of maltose itself, followed by hydrolysis (Fig. 22.10*b*), gives 2,3,4,6-tetra-*O*-methyl-D-glucose and 2,3,6-tri-*O*-methyl-D-glucose. The free —OH at C5 in the latter product indicates that it must have been involved in the oxide ring and that the reducing glucose is present as a *pyranose*.

22.12C Cellobiose

Partial hydrolysis of cellulose (Section 22.13C) gives the disaccharide cellobiose ($C_{12}H_{22}O_{11}$) (Fig. 22.11). Cellobiose resembles maltose in every respect except one: the configuration of its glycosidic linkage.

Cellobiose, like maltose, is a reducing sugar that, on acid-catalyzed hydrolysis, yields two molar equivalents of D-glucose. Cellobiose also undergoes mutarotation and forms a monophenylosazone. Methylation studies show that C1 of one glucose unit is connected

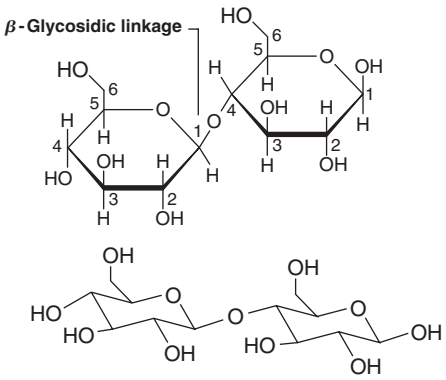
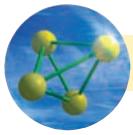


Figure 22.11 Two representations of the β anomer of cellobiose, 4-*O*-(β-D-glucopyranosyl)-β-D-glucopyranose.

in glycosidic linkage with C4 of the other and that both rings are six membered. Unlike maltose, however, cellobiose is hydrolyzed by *β-glucosidase* enzymes and not by α-glucosidase enzymes: This indicates that the glycosidic linkage in cellobiose is β (Fig. 22.11).



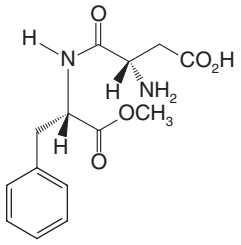
THE CHEMISTRY OF . . .

Artificial Sweeteners (How Sweet It Is)

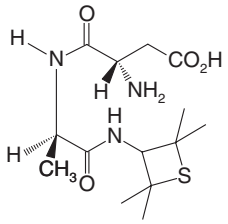
Sucrose (table sugar) and fructose are the most common natural sweeteners. We all know, however, that they add to our calorie intake and promote tooth decay. For these reasons, many people find artificial sweeteners to be an attractive alternative to the natural and calorie-contributing counterparts.

Some products that contain the artificial sweetener aspartame.

Perhaps the most successful and widely used artificial sweetener is aspartame, the methyl ester of a dipeptide formed from phenylalanine and aspartic acid (Section 24.3D). Aspartame is roughly 100 times as sweet as sucrose. It undergoes slow hydrolysis in solution, however, which limits its shelf life in products such as soft drinks. It also cannot be used for baking because it decomposes with heat. Furthermore, people with a genetic condition known as phenylketonuria cannot use aspartame because their metabolism causes a buildup of phenylpyruvic acid derived from aspartame. Accumulation of phenylpyruvic acid is harmful, especially to infants. Alitame, on the other hand, is a compound related to aspartame, but with improved properties. It is more stable than aspartame and roughly 2000 times as sweet as sucrose.

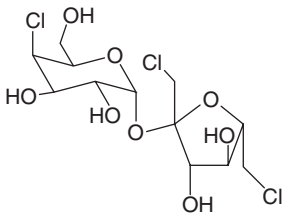


Aspartame



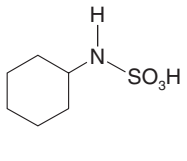
Alitame

Sucralose is a trichloro derivative of sucrose that is an artificial sweetener. Like aspartame, it is also approved for use by the U.S. Food and Drug Administration (FDA). Sucralose is 600 times sweeter than sucrose and has many properties desirable in an artificial sweetener. Sucralose looks and tastes like sugar, is stable at the temperatures used for cooking and baking, and it does not cause tooth decay or provide calories.

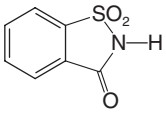


Sucralose

Cyclamate and saccharin, used as their sodium or calcium salts, were popular sweeteners at one time. A common formulation involved a 10:1 mixture of cyclamate and saccharin that proved sweeter than either compound individually. Tests showed, however, that this mixture produced tumors in animals, and the FDA subsequently banned it. Certain exclusions to the regulations nevertheless allow continued use of saccharin in some products.

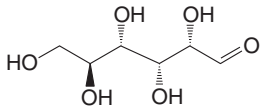


Cyclamate



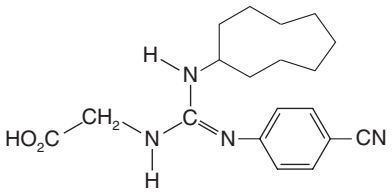
Saccharin

Many other compounds have potential as artificial sweeteners. For example, L sugars are also sweet, and they presumably would provide either zero or very few calories because our enzymes have evolved to selectively metabolize their enantiomers instead, the D sugars. Although sources of L sugars are rare in nature, all eight L-hexoses have been synthesized by S. Masamune and K. B. Sharpless using the Sharpless asymmetric epoxidation (Sections 11.13 and 22.11) and other enantioselective synthetic methods.



L-Glucose

Much of the research on sweeteners involves probing the structure of sweetness receptor sites. One model proposed for a sweetness receptor incorporates eight binding interactions that involve hydrogen bonding as well as van der Waals forces. Sucronic acid is a synthetic compound designed on the basis of this model. Sucronic acid is reported to be 200,000 times as sweet as sucrose.



Sucronic acid

22.12D Lactose

Lactose (Fig. 22.12) is a disaccharide present in the milk of humans, cows, and almost all other mammals. Lactose is a reducing sugar that hydrolyzes to yield D-glucose and D-galactose; the glycosidic linkage is β.

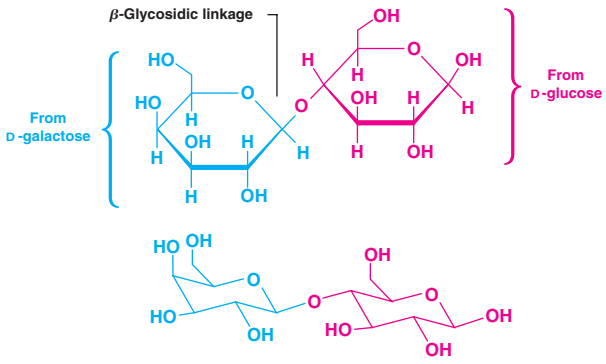


Figure 22.12 Two representations of the β anomer of lactose, 4-*O*-(β-D-galactopyranosyl)-β-D-glucopyranose.

22.13 Polysaccharides

- **Polysaccharides**, also known as **glycans**, consist of monosaccharides joined together by glycosidic linkages.

Polysaccharides that are polymers of a single monosaccharide are called **homopolysaccharides**; those made up of more than one type of monosaccharide are called **heteropolysaccharides**. Homopolysaccharides are also classified on the basis of their monosaccharide units. A homopolysaccharide consisting of glucose monomeric units is called a **glucan**; one consisting of galactose units is a **galactan**, and so on.

Three important polysaccharides, all of which are glucans, are starch, glycogen, and cellulose.

- Starch is the principal food reserve of plants, glycogen functions as a carbohydrate reserve for animals, and cellulose serves as structural material in plants.

As we examine the structures of these three polysaccharides, we shall be able to see how each is especially suited for its function.

22.13A Starch

Starch occurs as microscopic granules in the roots, tubers, and seeds of plants. Corn, potatoes, wheat, and rice are important commercial sources of starch. Heating starch with water causes the granules to swell and produce a colloidal suspension from which two major components can be isolated. One fraction is called *amylose* and the other *amylopectin*. Most starches yield 10–20% amylose and 80–90% amylopectin.

- Amylose typically consists of more than 1000 D-glucopyranoside units *connected in α linkages* between C1 of one unit and C4 of the next (Fig. 22.13).

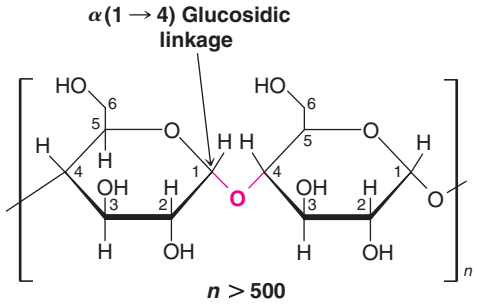


Figure 22.13 Partial structure of amylose, an unbranched polymer of D-glucose connected in α(1 → 4) glycosidic linkages.

Thus, in the ring size of its glucose units and in the configuration of the glycosidic linkages between them, amylose resembles maltose.

Chains of D-glucose units with α-glycosidic linkages such as those of amylose tend to assume a helical arrangement (Fig. 22.14). This arrangement results in a compact shape for the amylose molecule even though its molecular weight is quite large (150,000–600,000).

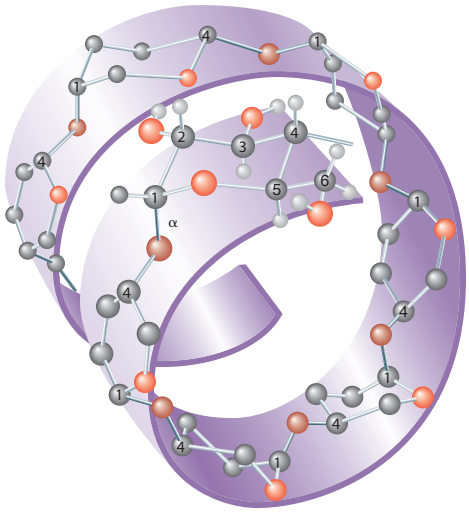


Figure 22.14 Amylose. The α(1 → 4) linkages cause it to assume the shape of a left-handed helix. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

- Amylopectin has a structure similar to that of amylose [i.e., α(1 → 4) links], except that in amylopectin the chains are branched. Branching takes place between C6 of one glucose unit and C1 of another and occurs at intervals of 20–25 glucose units (Fig. 22.15).

Physical measurements indicate that amylopectin has a molecular weight of 1–6 million; thus amylopectin consists of hundreds of interconnecting chains of 20–25 glucose units each.

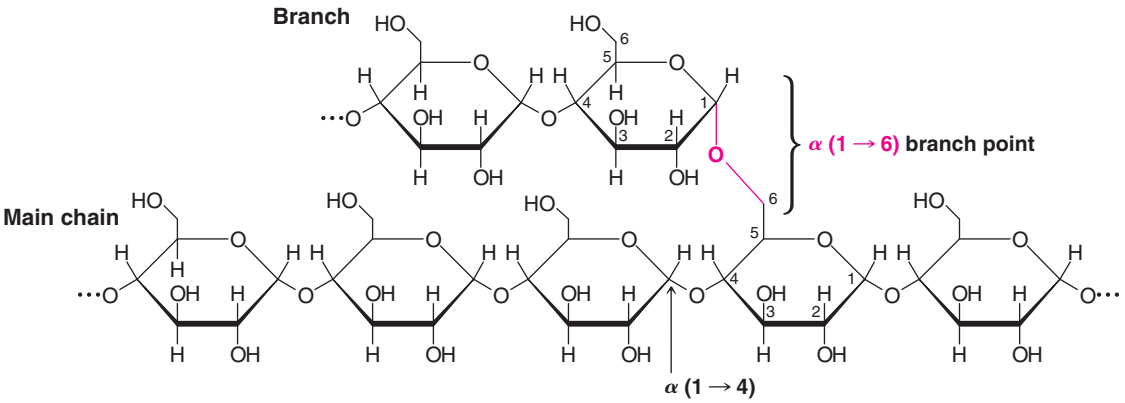


Figure 22.15 Partial structure of amylopectin.

22.13B Glycogen

- Glycogen has a structure very much like that of amylopectin; however, in glycogen the chains are much more highly branched.

Methylation and hydrolysis of glycogen indicate that there is one end group for every 10–12 glucose units; branches may occur as often as every 6 units. Glycogen has a very high molecular weight. Studies of glycogens isolated under conditions that minimize the likelihood of hydrolysis indicate molecular weights as high as 100 million.

The size and structure of glycogen beautifully suit its function as a reserve carbohydrate for animals. First, its size makes it too large to diffuse across cell membranes; thus, glycogen remains inside the cell, where it is needed as an energy source. Second, because glycogen incorporates tens of thousands of glucose units in a single molecule, it solves an important osmotic problem for the cell. Were so many glucose units present in the cell as individual molecules, the osmotic pressure within the cell would be enormous—so large that the cell membrane would almost certainly break.* Finally, the localization of glucose units within a large, highly branched structure simplifies one of the cell's logistical problems: that of having a ready source of glucose when cellular glucose concentrations are low and of being able to store glucose rapidly when cellular glucose concentrations are high. There are enzymes within the cell that catalyze the reactions by which glucose units are detached from (or attached to) glycogen. These enzymes operate at end groups by hydrolyzing (or forming) $\alpha(1 \rightarrow 4)$ glycosidic linkages. Because glycogen is so highly branched, a very large number of end groups is available at which these enzymes can operate. At the same time the overall concentration of glycogen (in moles per liter) is quite low because of its enormous molecular weight.

Amylopectin presumably serves a similar function in plants. The fact that amylopectin is less highly branched than glycogen is, however, not a serious disadvantage. Plants have a much lower metabolic rate than animals—and plants, of course, do not require sudden bursts of energy.

Animals store energy as fats (triacylglycerols) as well as glycogen. Fats, because they are more highly reduced, are capable of furnishing much more energy. The metabolism of a typical fatty acid, for example, liberates more than twice as much energy per carbon as glucose or glycogen. Why, then, we might ask, have two different energy repositories evolved? Glucose (from glycogen) is readily available and is highly water soluble.** Glucose, as a result, diffuses rapidly through the aqueous medium of the cell and serves as

*The phenomenon of osmotic pressure occurs whenever two solutions of different concentrations are separated by a membrane that allows penetration (by osmosis) of the solvent but not of the solute. The osmotic pressure (π) on one side of the membrane is related to the number of moles of solute particles (n), the volume of the solution (V), and the gas constant times the absolute temperature (RT): $\pi V = nRT$.

**Glucose is actually liberated as glucose-6-phosphate (G6P), which is also water soluble.

an ideal source of "ready energy." Long-chain fatty acids, by contrast, are almost insoluble in water, and their concentration inside the cell could never be very high. They would be a poor source of energy if the cell were in an energy pinch. On the other hand, fatty acids (as triacylglycerols), because of their caloric richness, are an excellent energy repository for long-term energy storage.

22.13C Cellulose

When we examine the structure of cellulose, we find another example of a polysaccharide in which nature has arranged monomeric glucose units in a manner that suits its function.

- Cellulose contains D-glucopyranoside units linked in (1 → 4) fashion in very long unbranched chains. Unlike starch and glycogen, however, the linkages in cellulose are *β-glycosidic linkages* (Fig. 22.16).

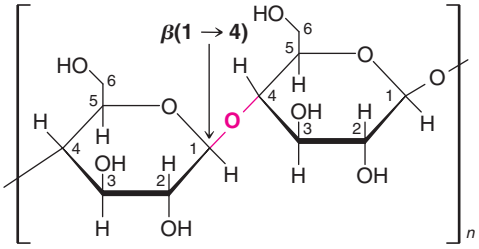


Figure 22.16 A portion of a cellulose chain. The glycosidic linkages are β(1 → 4).

The β-glycosidic linkages of cellulose make cellulose chains essentially linear; they do not tend to coil into helical structures as do glucose polymers when linked in an α(1 → 4) manner.

The linear arrangement of β-linked glucose units in cellulose presents a uniform distribution of —OH groups on the outside of each chain. When two or more cellulose chains make contact, the hydroxyl groups are ideally situated to "zip" the chains together by forming hydrogen bonds (Fig. 22.17). Zipping many cellulose chains together in this way gives a highly insoluble, rigid, and fibrous polymer that is ideal as cell-wall material for plants.

This special property of cellulose chains, we should emphasize, is not just a result of β(1 → 4) glycosidic linkages; it is also a consequence of the precise stereochemistry of D-glucose at each chirality center. Were D-galactose or D-allose units linked in a similar fashion, they almost certainly would not give rise to a polymer with properties like cellulose.

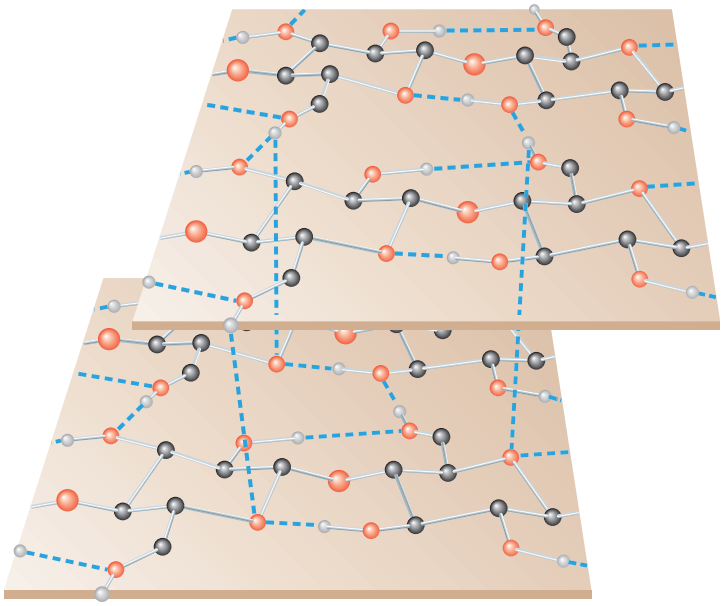


Figure 22.17 A proposed structure for cellulose. A fiber of cellulose may consist of about 40 parallel strands of glucose molecules linked in a β(1 → 4) fashion. Each glucose unit in a chain is turned over with respect to the preceding glucose unit and is held in this position by hydrogen bonds (dashed lines) between the chains. The glucan chains line up laterally to form sheets, and these sheets stack vertically so that they are staggered by one-half of a glucose unit. (Hydrogen atoms that do not participate in hydrogen bonding have been omitted for clarity.) (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

Thus, we get another glimpse of why D-glucose occupies such a special position in the chemistry of plants and animals. Not only is it the most stable aldohexose (because it can exist in a chair conformation that allows all of its bulky groups to occupy equatorial positions), but its special stereochemistry also allows it to form helical structures when α linked as in starches, and rigid linear structures when β linked as in cellulose.

There is another interesting and important fact about cellulose: The digestive enzymes of humans cannot attack its $\beta(1 \rightarrow 4)$ linkages. Hence, cellulose cannot serve as a food source for humans, as can starch. Cows and termites, however, can use cellulose (of grass and wood) as a food source because symbiotic bacteria in their digestive systems furnish β-glucosidase enzymes.

Perhaps we should ask ourselves one other question: Why has D-(+)-glucose been selected for its special role rather than L-(−)-glucose, its mirror image? Here an answer cannot be given with any certainty. The selection of D-(+)-glucose may simply have been a random event early in the course of the evolution of enzyme catalysts. Once this selection was made, however, the chirality of the active sites of the enzymes involved would retain a bias toward D-(+)-glucose and against L-(−)-glucose (because of the improper fit of the latter). Once introduced, this bias would be perpetuated and extended to other catalysts.

Finally, when we speak about evolutionary selection of a particular molecule for a given function, we do not mean to imply that evolution operates on a molecular level. Evolution, of course, takes place at the level of organism populations, and molecules are selected only in the sense that their use gives the organism an increased likelihood of surviving and procreating.

22.13D Cellulose Derivatives

A number of derivatives of cellulose are used commercially. Most of these are compounds in which two or three of the free hydroxyl groups of each glucose unit have been converted to an ester or an ether. This conversion substantially alters the physical properties of the material, making it more soluble in organic solvents and allowing it to be made into fibers and films. Treating cellulose with acetic anhydride produces the triacetate known as "Arnel" or "acetate," used widely in the textile industry. Cellulose trinitrate, also called "gun cotton" or nitrocellulose, is used in explosives.

Rayon is made by treating cellulose (from cotton or wood pulp) with carbon disulfide in a basic solution. This reaction converts cellulose to a soluble xanthate:

$$\text{Cellulose—OH} + CS_2 \xrightarrow{\text{NaOH}} \text{cellulose—O—C(=S)—S}^- \text{Na}^+$$

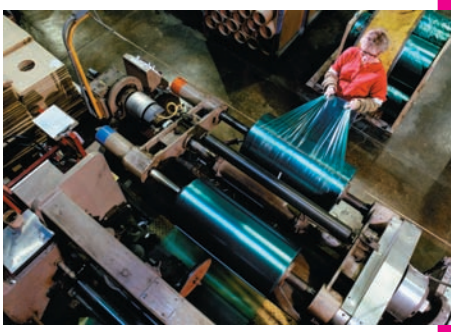
Cellulose xanthate

The solution of cellulose xanthate is then passed through a small orifice or slit into an acidic solution. This operation regenerates the —OH groups of cellulose, causing it to precipitate as a fiber or a sheet:

$$\text{Cellulose—O—C(=S)—S}^- \text{Na}^+ \xrightarrow{H_3O^+} \text{cellulose—OH}$$

Rayon or cellophane

The fibers are *rayon*; the sheets, after softening with glycerol, are *cellophane*.

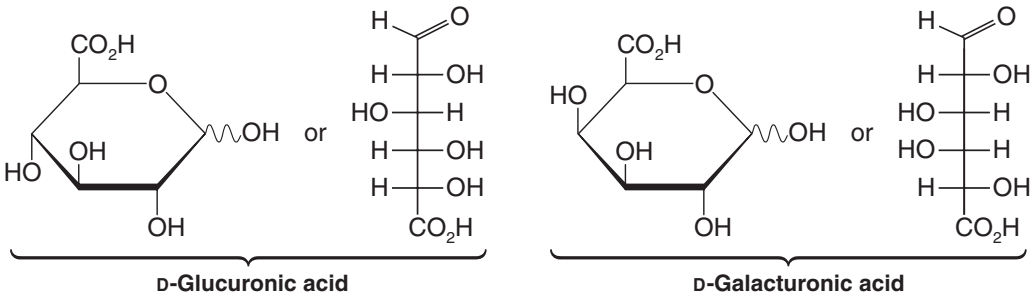


Cellophane on rollers at a manufacturing plant.

22.14 Other Biologically Important Sugars

Monosaccharide derivatives in which the —CH_2OH group at C6 has been specifically oxidized to a carboxyl group are called **uronic acids**. Their names are based on the monosaccharide from which they are derived. For example, specific oxidation of C6 of glucose

to a carboxyl group converts *glucose* to **glucuronic acid**. In the same way, specific oxidation of C6 of *galactose* would yield **galacturonic acid**:

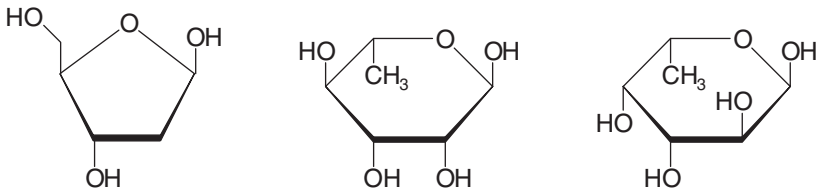


Review Problem 22.19

Direct oxidation of an aldose affects the aldehyde group first, converting it to a carboxylic acid (Section 22.6B), and most oxidizing agents that will attack 1° alcohol groups will also attack 2° alcohol groups. Clearly, then, a laboratory synthesis of a uronic acid from an aldose requires protecting these groups from oxidation. Keeping this in mind, suggest a method for carrying out a specific oxidation that would convert D-galactose to D-galacturonic acid. (*Hint*: See Section 22.5E.)

- Monosaccharides in which an —OH group has been replaced by —H are known as **deoxy sugars**.

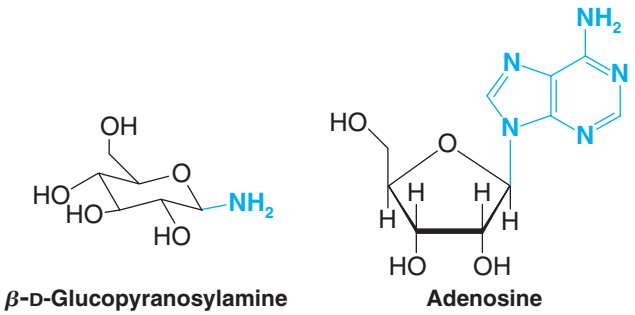
The most important deoxy sugar, because it occurs in DNA, is **deoxyribose**. Other deoxy sugars that occur widely in polysaccharides are L-rhamnose and L-fucose:



22.15 Sugars That Contain Nitrogen

22.15A Glycosylamines

A sugar in which an amino group replaces the anomeric —OH is called a glycosylamine. Examples are β-D-glucopyranosylamine and adenosine:



Adenosine is an example of a glycosylamine that is also called a nucleoside.

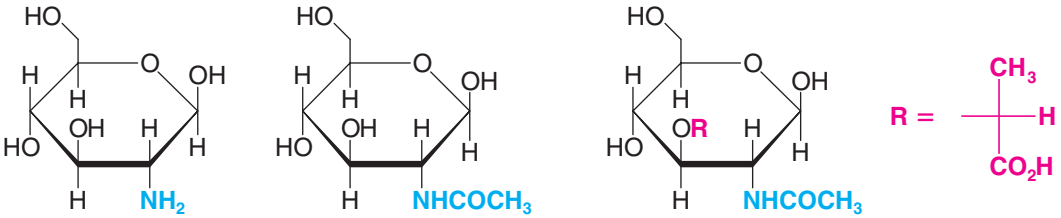
- **Nucleosides** are glycosylamines in which the amino component is a pyrimidine or a purine (Section 20.1B) and in which the sugar component is either D-ribose or 2-deoxy-D-ribose (i.e., D-ribose minus the oxygen at the 2 position).

Nucleosides are the important components of RNA (ribonucleic acid) and DNA (deoxyribonucleic acid). We shall describe their properties in detail in Section 25.2.

22.15B Amino Sugars

- A sugar in which an amino group replaces a nonanomeric —OH group is called an **amino sugar**.

D-Glucosamine is an example of an amino sugar. In many instances the amino group is acetylated as in ***N*-acetyl-D-glucosamine**. ***N*-Acetylmuramic acid** is an important component of bacterial cell walls (Section 24.10).



D-Glucosamine can be obtained by hydrolysis of **chitin**, a polysaccharide found in the shells of lobsters and crabs and in the external skeletons of insects and spiders. The amino group of D-glucosamine as it occurs in chitin, however, is acetylated; thus, the repeating unit is actually *N*-acetylglucosamine (Fig. 22.18). The glycosidic linkages in chitin are $\beta(1 \rightarrow 4)$. X-Ray analysis indicates that the structure of chitin is similar to that of cellulose.

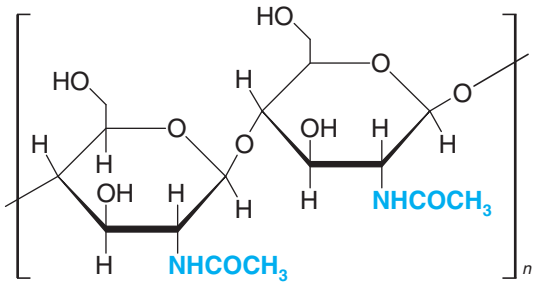


Figure 22.18 A partial structure of chitin. The repeating units are *N*-acetylglucosamines linked $\beta(1 \rightarrow 4)$.

D-Glucosamine can also be isolated from **heparin**, a sulfated polysaccharide that consists predominately of alternating units of D-glucuronate-2-sulfate and *N*-sulfo-D-glucosamine-6-sulfate (Fig. 22.19). Heparin occurs in intracellular granules of mast cells that line arterial walls, where, when released through injury, it inhibits the clotting of blood. Its purpose seems to be to prevent runaway clot formation. Heparin is widely used in medicine to prevent blood clotting in postsurgical patients.

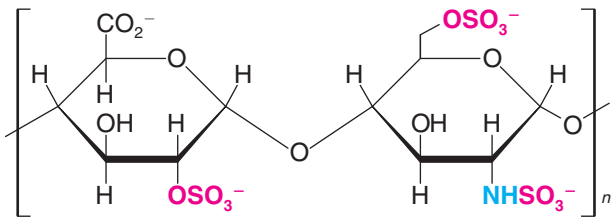
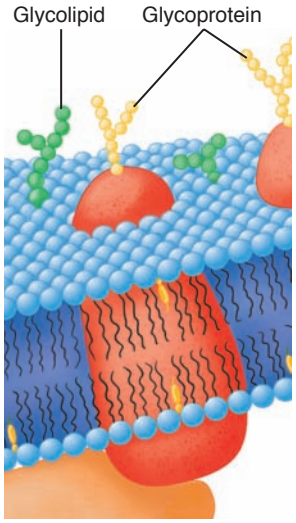


Figure 22.19 A partial structure of heparin, a polysaccharide that prevents blood clotting.

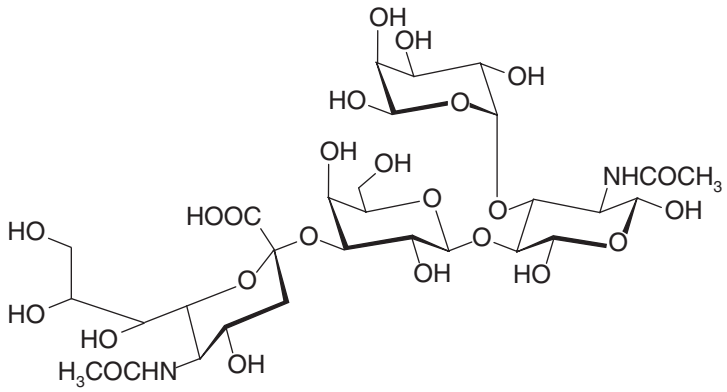
22.16 *Glycolipids and Glycoproteins of the Cell Surface: Cell Recognition and the Immune System*



Reprinted with permission of John Wiley & Sons, Inc., from Voet, D., and Voet, J.G. *Biochemistry*, Second Edition © 1995 Voet, D., and Voet, J.G.

Before 1960, it was thought that the biology of carbohydrates was rather uninteresting, that, in addition to being a kind of inert filler in cells, carbohydrates served only as an energy source and, in plants, as structural materials. Research has shown, however, that carbohydrates joined through glycosidic linkages to lipids (Chapter 23) and to proteins (Chapter 24), called **glycolipids** and **glycoproteins**, respectively, have functions that span the entire spectrum of activities in the cell. Indeed, most proteins are glycoproteins, of which the carbohydrate content can vary from less than 1% to greater than 90%.

Glycolipids and glycoproteins on the cell surface (Section 23.6A) are now known to be the agents by which cells interact with other cells and with invading bacteria and viruses. The immune system's role in healing and in autoimmune diseases such as rheumatoid arthritis involves cell recognition through cell surface carbohydrates. Important carbohydrates in this role are sialyl Lewis^x acids (see the chapter opening vignette). Tumor cells also have specific carbohydrate markers on their surface as well, a fact that may make it possible to develop vaccines against cancer. (See "The Chemistry of . . . Vaccines Against Cancer" in *WileyPLUS*.)



A sialyl Lewis^x acid

The human blood groups offer another example of how carbohydrates, in the form of glycolipids and glycoproteins, act as biochemical markers. The A, B, and O blood types are determined, respectively, by the A, B, and H determinants on the blood cell surface. (The odd naming of the type O determinant came about for complicated historical reasons.) Type AB blood cells have both A and B determinants. These determinants are the carbohydrate portions of the A, B, and H **antigens**.

Antigens are characteristic chemical substances that cause the production of **antibodies** when injected into an animal. Each antibody can bind at least two of its corresponding antigen molecules, causing them to become linked. Linking of red blood cells causes them to agglutinate (clump together). In a transfusion this agglutination can lead to a fatal blockage of the blood vessels.

Individuals with type A antigens on their blood cells carry anti-B antibodies in their serum; those with type B antigens on their blood cells carry anti-A antibodies in their serum. Individuals with type AB cells have both A and B antigens but have neither anti-A nor anti-B antibodies. Type O individuals have neither A nor B antigens on their blood cells but have both anti-A and anti-B antibodies.

Helpful Hint

See "The Chemistry of... Oligosaccharide Synthesis on a Solid Support–the Glycal Assembly Approach" in *WileyPLUS* regarding the synthesis of promising carbohydrate anticancer vaccines.

The A, B, and H antigens differ only in the monosaccharide units at their nonreducing ends. The type H antigen (Fig. 22.20) is the precursor oligosaccharide of the type A and B antigens. Individuals with blood type A have an enzyme that specifically adds an *N*-acetylgalactosamine unit to the 3-OH group of the terminal galactose unit of the H antigen. Individuals with blood type B have an enzyme that specifically adds galactose instead. In individuals with type O blood, the enzyme is inactive.

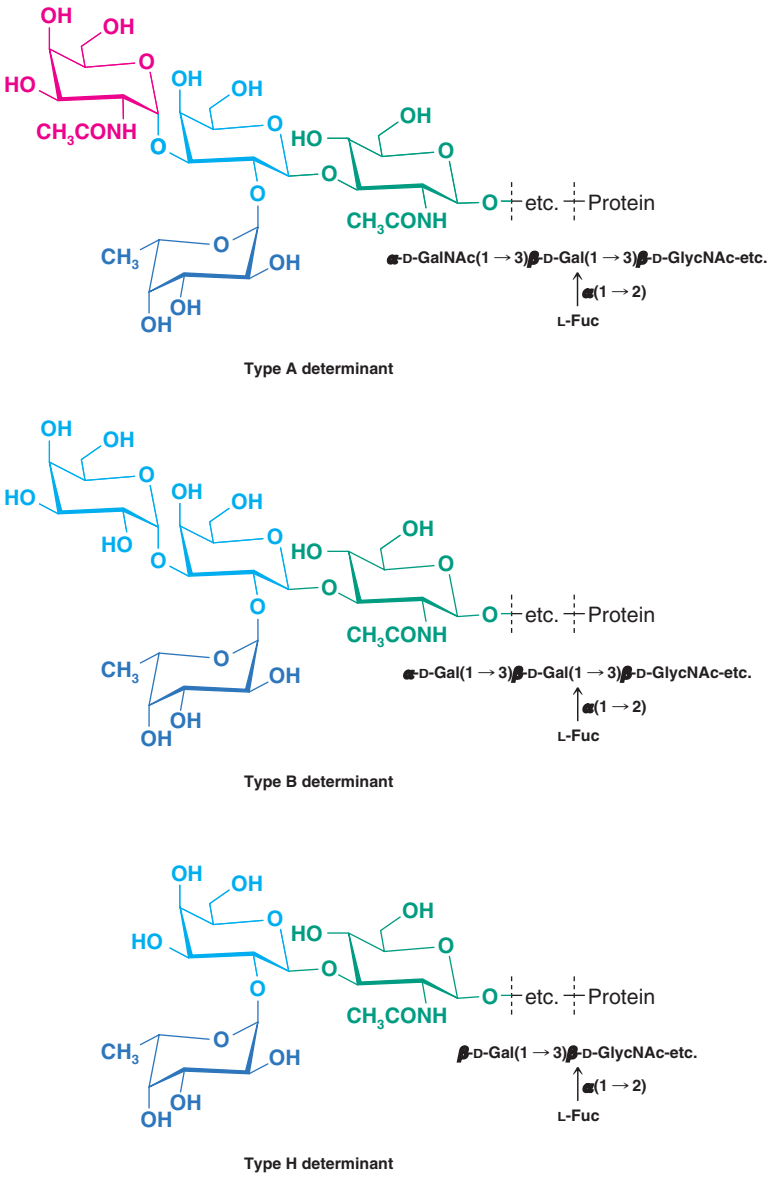
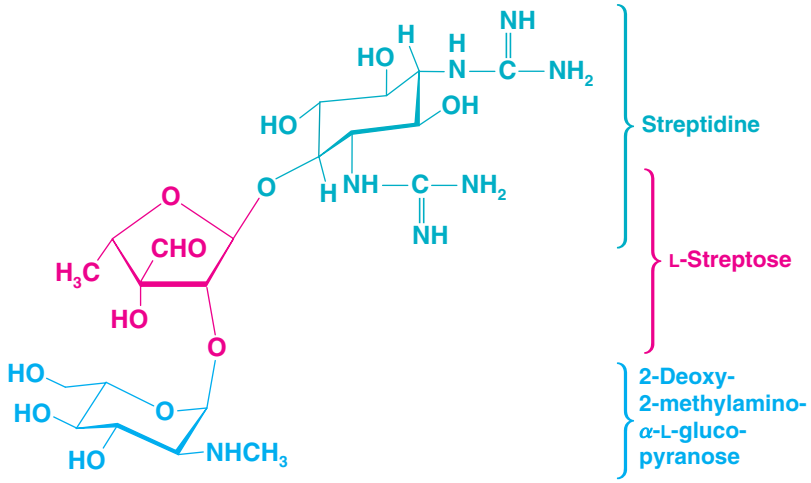


Figure 22.20 The terminal monosaccharides of the antigenic determinants for types A, B, and O blood. The type H determinant is present in individuals with blood type O and is the precursor of the type A and B determinants. These oligosaccharide antigens are attached to carrier lipid or protein molecules that are anchored in the red blood cell membrane (see Fig. 23.9 for a depiction of a cell membrane). Ac = acetyl, Gal = D-galactose, GalNAc = *N*-acetylgalactosamine, GlycNAc = *N*-acetylglucosamine, Fuc = fucose.

Antigen–antibody interactions like those that determine blood types are the basis of the immune system. These interactions often involve the chemical recognition of a glycolipid or glycoprotein in the antigen by a glycolipid or glycoprotein of the antibody. In "The Chemistry of . . . Antibody-Catalyzed Aldol Condensations" (in *WileyPLUS*, Chapter 19), however, we saw a different and emerging dimension of chemistry involving antibodies. We shall explore this topic further in the Chapter 24 opening vignette on designer catalysts and in "The Chemistry of . . . Some Catalytic Antibodies" (Section 24.12).

22.17 Carbohydrate Antibiotics

One of the important discoveries in carbohydrate chemistry was the isolation (in 1944) of the carbohydrate antibiotic called *streptomycin*. Streptomycin disrupts bacterial protein synthesis. Its structure is made up of the following three subunits:



All three components are unusual: The amino sugar is based on L-glucose; streptose is a branched-chain monosaccharide; and streptidine is not a sugar at all, but a cyclohexane derivative called an amino cyclitol.

Other members of this family are antibiotics called kanamycins, neomycins, and gentamicins (not shown). All are based on an amino cyclitol linked to one or more amino sugars. The glycosidic linkage is nearly always α. These antibiotics are especially useful against bacteria that are resistant to penicillins.

22.18 Summary of Reactions of Carbohydrates

The reactions of carbohydrates, with few exceptions, are the reactions of functional groups that we have studied in earlier chapters, especially those of aldehydes, ketones, and alcohols. The most central reactions of carbohydrates are those of hemiacetal and acetal formation and hydrolysis. Hemiacetal groups form the pyranose and furanose rings in carbohydrates, and acetal groups form glycoside derivatives and join monosaccharides together to form di-, tri-, oligo-, and polysaccharides.

Other reactions of carbohydrates include those of alcohols, carboxylic acids, and their derivatives. Alkylation of carbohydrate hydroxyl groups leads to ethers. Acylation of their hydroxyl groups produces esters. Alkylation and acylation reactions are sometimes used to protect carbohydrate hydroxyl groups from reaction while a transformation occurs elsewhere. Hydrolysis reactions are involved in converting ester and lactone derivatives of carbohydrates back to their polyhydroxy form. Enolization of aldehydes and ketones leads to epimerization and interconversion of aldoses and ketoses. Addition reactions of aldehydes and ketones are useful, too, such as the addition of ammonia derivatives in osazone formation, and of cyanide in the Kiliani–Fischer synthesis. Hydrolysis of nitriles from the Kiliani–Fischer synthesis leads to carboxylic acids.

Oxidation and reduction reactions have their place in carbohydrate chemistry as well. Reduction reactions of aldehydes and ketones, such as borohydride reduction and catalytic hydrogenation, are used to convert aldoses and ketoses to alditols. Oxidation by Tollens' and Benedict's reagents is a test for the hemiacetal linkage in a sugar. Bromine water oxidizes the aldehyde group of an aldose to an aldonic acid. Nitric acid oxidizes both the aldehyde group and terminal hydroxymethyl group of an aldose to an aldaric acid (a dicarboxylic acid). Lastly, periodate cleavage of carbohydrates yields oxidized fragments that can be useful for structure elucidation.

Key Terms and Concepts

The key terms and concepts that are highlighted in **bold, blue text** within the chapter are defined in the glossary (at the back of the book) and have hyperlinked definitions in the accompanying ***WileyPLUS*** course (www.wileyplus.com).

Problems

Note to Instructors: Many of the homework problems are available for assignment via *Wiley PLUS*, an online teaching and learning solution.

CARBOHYDRATE STRUCTURE AND REACTIONS

22.20 Give appropriate structural formulas to illustrate each of the following:

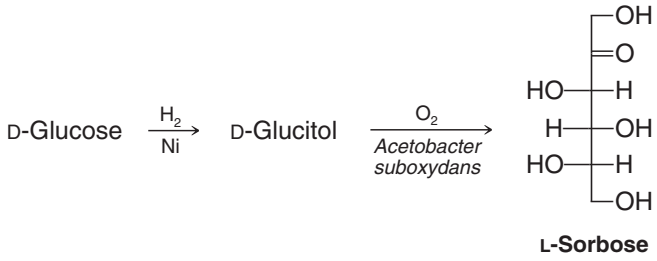
(a) An aldopentose
(b) A ketohexose
(c) An L-monosaccharide
(d) A glycoside
(e) An aldonic acid
(f) An aldaric acid
(g) An aldonolactone
(h) A pyranose
(i) A furanose
(j) A reducing sugar
(k) A pyranoside
(l) A furanoside
(m) Epimers
(n) Anomers
(o) A phenylosazone
(p) A disaccharide
(q) A polysaccharide
(r) A nonreducing sugar

22.21 Draw conformational formulas for each of the following: **(a)** α-D-allopyranose, **(b)** methyl β-D-allopyranoside, and **(c)** methyl 2,3,4,6-tetra-*O*-methyl-β-D-allopyranoside.

22.22 Draw structures for furanose and pyranose forms of D-ribose. Show how you could use periodate oxidation to distinguish between a methyl ribofuranoside and a methyl ribopyranoside.

22.23 One reference book lists D-mannose as being dextrorotatory; another lists it as being levorotatory. Both references are correct. Explain.

22.24 The starting material for a commercial synthesis of vitamin C is L-sorbose (see the following reaction); it can be synthesized from D-glucose through the following reaction sequence:



L-Sorbose

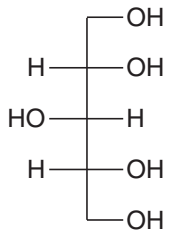
The second step of this sequence illustrates the use of a bacterial oxidation; the microorganism *A. suboxydans* accomplishes this step in 90% yield. The overall result of the synthesis is the transformation of a D-aldohexose (D-glucose) into an L-ketohexose (L-sorbose). What does this mean about the specificity of the bacterial oxidation?

22.25 What two aldoses would yield the same phenylosazone as L-sorbose (Problem 22.24)?

22.26 In addition to fructose (Review Problem 22.12) and sorbose (Problem 22.24), there are two other 2-ketohexoses, *psicose* and *tagatose*. D-Psicose yields the same phenylosazone as D-allose (or D-altrose); D-tagatose yields the same osazone as D-galactose (or D-talose). What are the structures of D-psicose and D-tagatose?

22.27 **A, B**, and **C** are three aldohexoses. Compounds **A** and **B** yield the same optically active alditol when they are reduced with hydrogen and a catalyst; **A** and **B** yield different phenylosazones when treated with phenylhydrazine; **B** and **C** give the same phenylosazone but different alditols. Assuming that all are D sugars, give names and structures for **A, B**, and **C**.

22.28 Xylitol is a sweetener that is used in sugarless chewing gum. Starting with an appropriate monosaccharide, outline a possible synthesis of xylitol.



Xylitol

22.29 Although monosaccharides undergo complex isomerizations in base (see Section 22.5A), aldonic acids are epimerized specifically at C2 when they are heated with pyridine. Show how you could make use of this reaction in a synthesis of D-mannose from D-glucose.

22.30 The most stable conformation of most aldopyranoses is one in which the largest group, the $-CH_2OH$ group, is equatorial. However, D-idopyranose exists primarily in a conformation with an axial $-CH_2OH$ group. Write formulas for the two chair conformations of α-D-idopyranose (one with the $-CH_2OH$ group axial and one with the $-CH_2OH$ group equatorial) and provide an explanation.

STRUCTURE ELUCIDATION

22.31 **(a)** Heating D-altrose with dilute acid produces a nonreducing *anhydro sugar* ($C_6H_{10}O_5$). Methylation of the anhydro sugar followed by acid hydrolysis yields 2,3,4-tri-*O*-methyl-D-altrose. The formation of the anhydro sugar takes place through a chair conformation of β-D-altropyranose in which the $-CH_2OH$ group is axial. What is the structure of the anhydro sugar, and how is it formed? **(b)** D-Glucose also forms an anhydro sugar but the conditions required are much more drastic than for the corresponding reaction of D-altrose. Explain.

22.32 Show how the following experimental evidence can be used to deduce the structure of lactose (Section 22.12D):

1. Acid hydrolysis of lactose ($C_{12}H_{22}O_{11}$) gives equimolar quantities of D-glucose and D-galactose. Lactose undergoes a similar hydrolysis in the presence of a *β-galactosidase*.
2. Lactose is a reducing sugar and forms a phenylosazone; it also undergoes mutarotation.
3. Oxidation of lactose with bromine water followed by hydrolysis with dilute acid gives D-galactose and D-gluconic acid.
4. Bromine water oxidation of lactose followed by methylation and hydrolysis gives 2,3,6-tri-*O*-methylgluconolactone and 2,3,4,6-tetra-*O*-methyl-D-galactose.
5. Methylation and hydrolysis of lactose give 2,3,6-tri-*O*-methyl-D-glucose and 2,3,4,6-tetra-*O*-methyl-D-galactose.

22.33 Deduce the structure of the disaccharide *melibiose* from the following data:

1. Melibiose is a reducing sugar that undergoes mutarotation and forms a phenylosazone.
2. Hydrolysis of melibiose with acid or with an *α-galactosidase* gives D-galactose and D-glucose.
3. Bromine water oxidation of melibiose gives *melibionic acid*. Hydrolysis of melibionic acid gives D-galactose and D-gluconic acid. Methylation of melibionic acid followed by hydrolysis gives 2,3,4,6-tetra-*O*-methyl-D-galactose and 2,3,4,5-tetra-*O*-methyl-D-gluconic acid.
4. Methylation and hydrolysis of melibiose give 2,3,4,6-tetra-*O*-methyl-D-galactose and 2,3,4-tri-*O*-methyl-D-glucose.

22.34 Trehalose is a disaccharide that can be obtained from yeasts, fungi, sea urchins, algae, and insects. Deduce the structure of trehalose from the following information:

1. Acid hydrolysis of trehalose yields only D-glucose.
2. Trehalose is hydrolyzed by α-glucosidase but not by β-glucosidase enzymes.
3. Trehalose is a nonreducing sugar; it does not mutarotate, form a phenylosazone, or react with bromine water.
4. Methylation of trehalose followed by hydrolysis yields two molar equivalents of 2,3,4,6-tetra-*O*-methyl-D-glucose.

22.35 Outline chemical tests that will distinguish between members of each of the following pairs:

(a) D-Glucose and D-glucitol

(b) D-Glucitol and D-glucaric acid

(c) D-Glucose and D-fructose

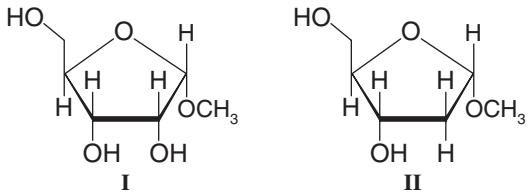
(d) D-Glucose and D-galactose

(e) Sucrose and maltose

(f) Maltose and maltonic acid

(g) Methyl β-D-glucopyranoside and 2,3,4,6-tetra-*O*-methyl-β-D-glucopyranose

(h) Methyl α-D-ribofuranoside (**I**) and methyl 2-deoxy-α-D-ribofuranoside (**II**):



22.36 A group of oligosaccharides called *Schardinger dextrins* can be isolated from *Bacillus macerans* when the bacillus is grown on a medium rich in amylose. These oligosaccharides are all *nonreducing*. A typical Schardinger dextrin undergoes hydrolysis when treated with an acid or an α-glucosidase to yield six, seven, or eight molecules of D-glucose. Complete methylation of a Schardinger dextrin followed by acid hydrolysis yields only 2,3,6-tri-*O*-methyl-D-glucose. Propose a general structure for a Schardinger dextrin.

22.37 *Isomaltose* is a disaccharide that can be obtained by enzymatic hydrolysis of amylopectin. Deduce the structure of isomaltose from the following data:

1. Hydrolysis of 1 mol of isomaltose by acid or by an α-glucosidase gives 2 mol of D-glucose.
2. Isomaltose is a reducing sugar.
3. Isomaltose is oxidized by bromine water to isomaltonic acid. Methylation of isomaltonic acid and subsequent hydrolysis yields 2,3,4,6-tetra-*O*-methyl-D-glucose and 2,3,4,5-tetra-*O*-methyl-D-gluconic acid.
4. Methylation of isomaltose itself followed by hydrolysis gives 2,3,4,6-tetra-*O*-methyl-D-glucose and 2,3,4-tri-*O*-methyl-D-glucose.

22.38 *Stachyose* occurs in the roots of several species of plants. Deduce the structure of stachyose from the following data:

1. Acidic hydrolysis of 1 mol of stachyose yields 2 mol of D-galactose, 1 mol of D-glucose, and 1 mol of D-fructose.
2. Stachyose is a nonreducing sugar.
3. Treating stachyose with an α-galactosidase produces a mixture containing D-galactose, sucrose, and a nonreducing trisaccharide called *raffinose*.
4. Acidic hydrolysis of raffinose gives D-glucose, D-fructose, and D-galactose. Treating raffinose with an α-galactosidase yields D-galactose and sucrose. Treating raffinose with invertase (an enzyme that hydrolyzes sucrose) yields fructose and *melibiose* (see Problem 22.33).
5. Methylation of stachyose followed by hydrolysis yields 2,3,4,6-tetra-*O*-methyl-D-galactose, 2,3,4-tri-*O*-methyl-D-galactose, 2,3,4-tri-*O*-methyl-D-glucose, and 1,3,4,6-tetra-*O*-methyl-D-fructose.

SPECTROSCOPY

22.39 *Arbutin*, a compound that can be isolated from the leaves of barberry, cranberry, and pear trees, has the molecular formula $C_{12}H_{16}O_7$. When arbutin is treated with aqueous acid or with a β-glucosidase, the reaction produces D-glucose and a compound **X** with the molecular formula $C_6H_6O_2$. The ^1H NMR spectrum of compound **X** consists of two singlets, one at δ 6.8 (4H) and one at δ 7.9 (2H). Methylation of arbutin followed by acidic hydrolysis yields 2,3,4,6-tetra-*O*-methyl-D-glucose and a compound **Y** ($C_7H_8O_2$). Compound **Y** is soluble in dilute aqueous NaOH but is insoluble in aqueous $NaHCO_3$. The ^1H NMR spectrum of **Y** shows a singlet at δ 3.9 (3H), a singlet at δ 4.8 (1H), and a multiplet (that resembles a singlet) at δ 6.8 (4H). Treating compound **Y** with aqueous NaOH and $(CH_3)_2SO_4$ produces compound **Z** ($C_8H_{10}O_2$). The ^1H NMR spectrum of **Z** consists of two singlets, one at δ 3.75 (6H) and one at δ 6.8 (4H). Propose structures for arbutin and for compounds **X, Y**, and **Z**.

22.40 When subjected to a Ruff degradation, a D-aldopentose, **A**, is converted to an aldotetrose, **B**. When reduced with sodium borohydride, the aldotetrose **B** forms an optically active alditol. The ^{13}C NMR spectrum of this alditol displays only two signals. The alditol obtained by direct reduction of **A** with sodium borohydride is not optically active. When **A** is used as the starting material for a Kiliani–Fischer synthesis, two diastereomeric aldohexoses, **C** and **D**, are produced. On treatment with sodium borohydride, **C** leads to an alditol **E**, and **D** leads to **F**. The ^{13}C NMR spectrum of **E** consists of three signals; that of **F** consists of six. Propose structures for A–F.

22.41 Figure 22.21 shows the ^{13}C NMR spectrum for the product of the reaction of D-(+)-mannose with acetone containing a trace of acid. This compound is a mannofuranose with some hydroxyl groups protected as acetone acetals (as acetonides). Use the ^{13}C NMR spectrum to determine how many acetonide groups are present in the compound.

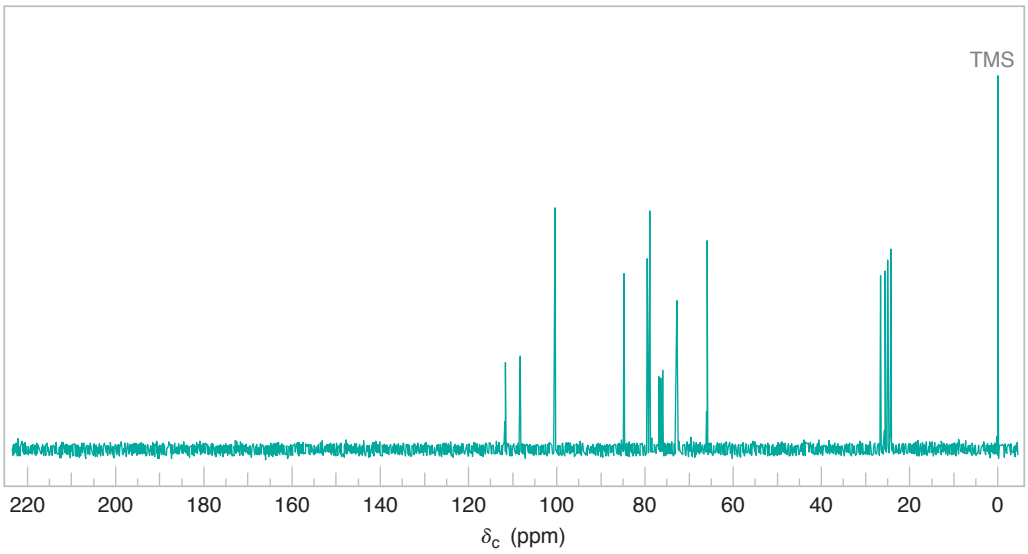
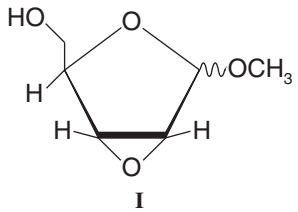


Figure 22.21 The broadband proton-decoupled ^{13}C NMR spectrum for the reaction product in Problem 22.41.

22.42 D-(+)-Mannose can be reduced with sodium borohydride to form D-mannitol. When D-mannitol is dissolved in acetone containing a trace amount of acid and the product of this reaction subsequently oxidized with $NaIO_4$, a compound whose ^{13}C NMR spectrum consists of six signals is produced. One of these signals is near δ 200. What is the structure of this compound?

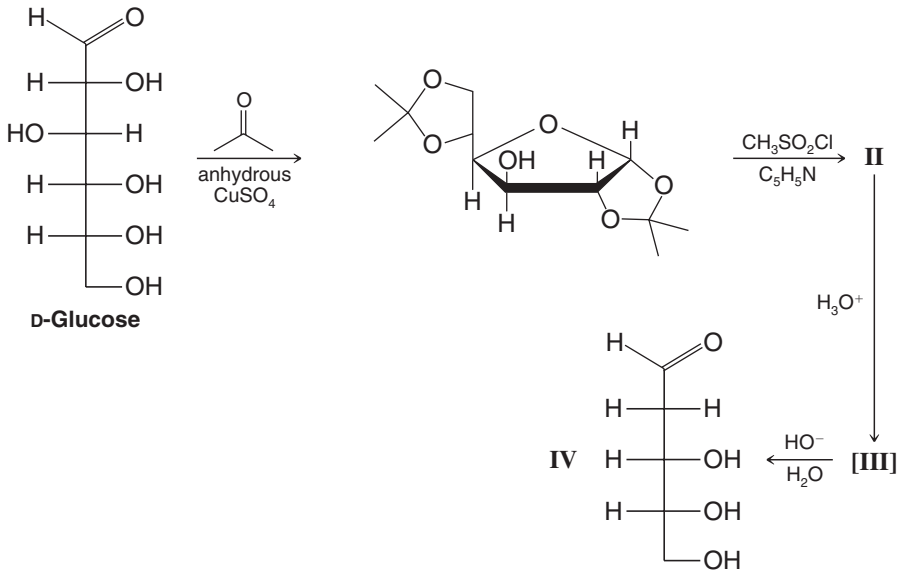
Challenge Problems

22.43 Of the two anomers of methyl 2,3-anhydro-D-ribofuranoside, **I**, the β form has a strikingly lower boiling point. Suggest an explanation using their structural formulas.



I

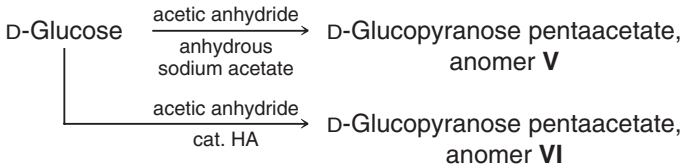
22.44 The following reaction sequence represents an elegant method of synthesis of 2-deoxy-D-ribose, **IV**, published by D. C. C. Smith in 1955:



(a) What are the structures of **II** and **III**?

(b) Propose a mechanism for the conversion of **III** to **IV**.

22.45



The ^{1}H NMR data for the two anomers included very comparable peaks in the δ 2.0–5.6 region but differed in that, as their highest δ peaks, anomer **V** had a doublet at δ 5.8 (1H, J = 12 Hz) while anomer **VI** had a doublet at δ 6.3 (1H, J = 4 Hz).

(a) Which proton in these anomers would be expected to have these highest δ values?

(b) Why do the signals for these protons appear as doublets?

(c) The relationship between the magnitude of the observed coupling constant and the dihedral angle (when measured using a Newman projection) between C—H bonds on the adjacent carbons of a C—C bond is given by the Karplus equation. It indicates that an axial–axial relationship results in a coupling constant of about 9 Hz (observed range is 8–14 Hz) and an equatorial–axial relationship results in a coupling constant of about 2 Hz (observed range is 1–7 Hz). Which of **V** and **VI** is the α anomer and which is the β anomer?

(d) Draw the most stable conformer for each of **V** and **VI**.

Learning Group Problems

1. **(a)** The members of one class of low-calorie sweeteners are called polyols. The chemical synthesis of one such polyol sweetener involves reduction of a certain disaccharide to a mixture of diastereomeric glycosides. The alcohol (actually polyol) portion of the diastereomeric glycosides derives from one of the sugar moieties in the original disaccharide. Exhaustive methylation of the sweetener (e.g., with dimethyl sulfate in the presence of hydroxide) followed by hydrolysis would be expected to produce 2,3,4,6-tetra-*O*-methyl-α-D-glucopyranose, 1,2,3,4,5-penta-*O*-methyl-D-sorbitol, and 1,2,3,4,5-penta-*O*-methyl-D-mannitol, in the ratio of 2:1:1. On the basis of this information, deduce the structure of the two disaccharide glycosides that make up the diastereomeric mixture in this polyol sweetener.

(b) Knowing that the mixture of two disaccharide glycosides in this sweetener results from reduction of a single disaccharide starting material (e.g., reduction by sodium borohydride), what would be the structure of the disaccharide *reactant* for the reduction step? Explain how reduction of this compound would produce the two glycosides.

(c) Write the lowest energy chair conformational structure for 2,3,4,6-tetra-*O*-methyl-α-D-glucopyranose.

2. Shikimic acid is a key biosynthetic intermediate in plants and microorganisms. In nature, shikimic acid is converted to chorismate, which is then converted to prephenate, ultimately leading to aromatic amino acids and other essential plant and microbial metabolites (see the Chapter 21 Learning Group problem). In the course of research on biosynthetic pathways involving shikimic acid, H. Floss (University of Washington) required shikimic acid labeled with ^{13}C to trace the destiny of the labeled carbon atoms in later biochemical transformations. To synthesize the labeled shikimic acid, Floss adapted a synthesis of optically active shikimic acid from D-mannose reported earlier by G. W. J. Fleet (Oxford University). This synthesis is a prime example of how natural sugars can be excellent chiral starting materials for the chemical synthesis of optically active target molecules. It is also an excellent example of classic reactions in carbohydrate chemistry. The Fleet–Floss synthesis of D-(−)-[1,7-^{13}C]-shikimic acid **(1)** from D-mannose is shown in Scheme 1.

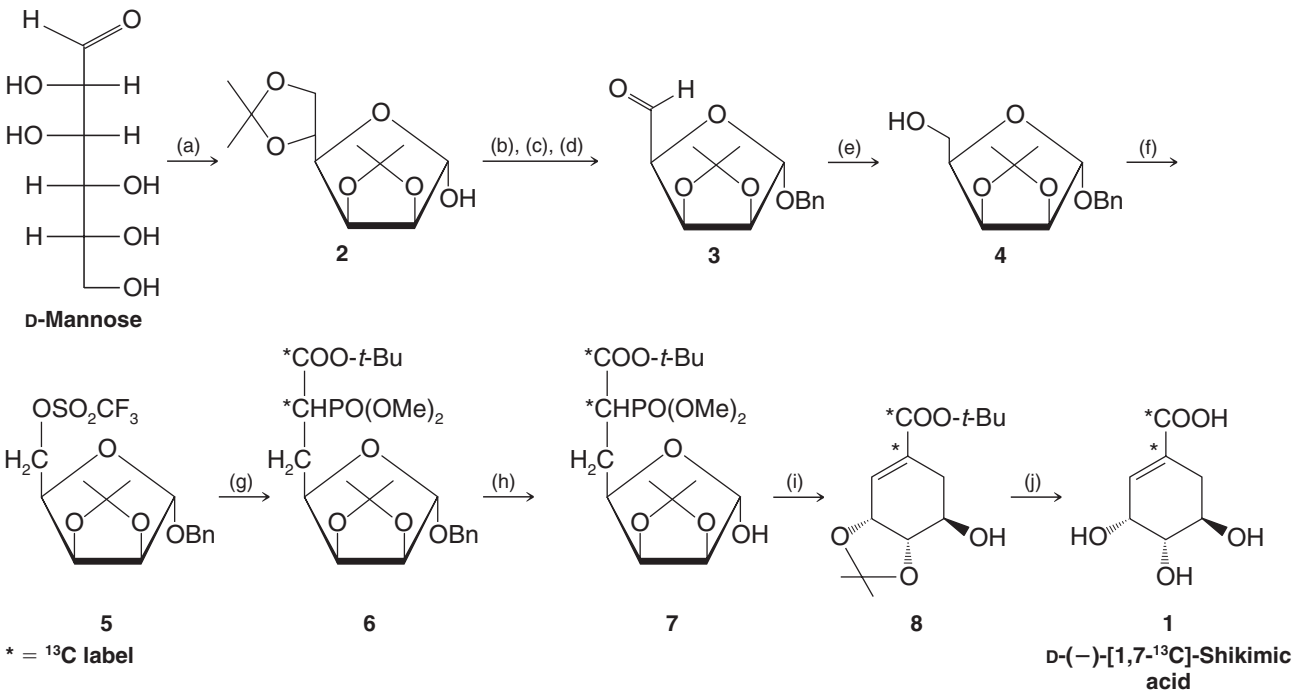
(a) Comment on the several transformations that occur between D-mannose and **2**. What new functional groups are formed?

(b) What is accomplished in the steps from **2** to **3, 3** to **4**, and **4** to **5**?

(c) Deduce the structure of compound **9** (a reagent used to convert **5** to **6**), knowing that it was a carbanion that displaced the trifluoromethanesulfonate (triflate) group of **5**. Note that it was compound **9** that brought with it the required ^{13}C atoms for the final product.

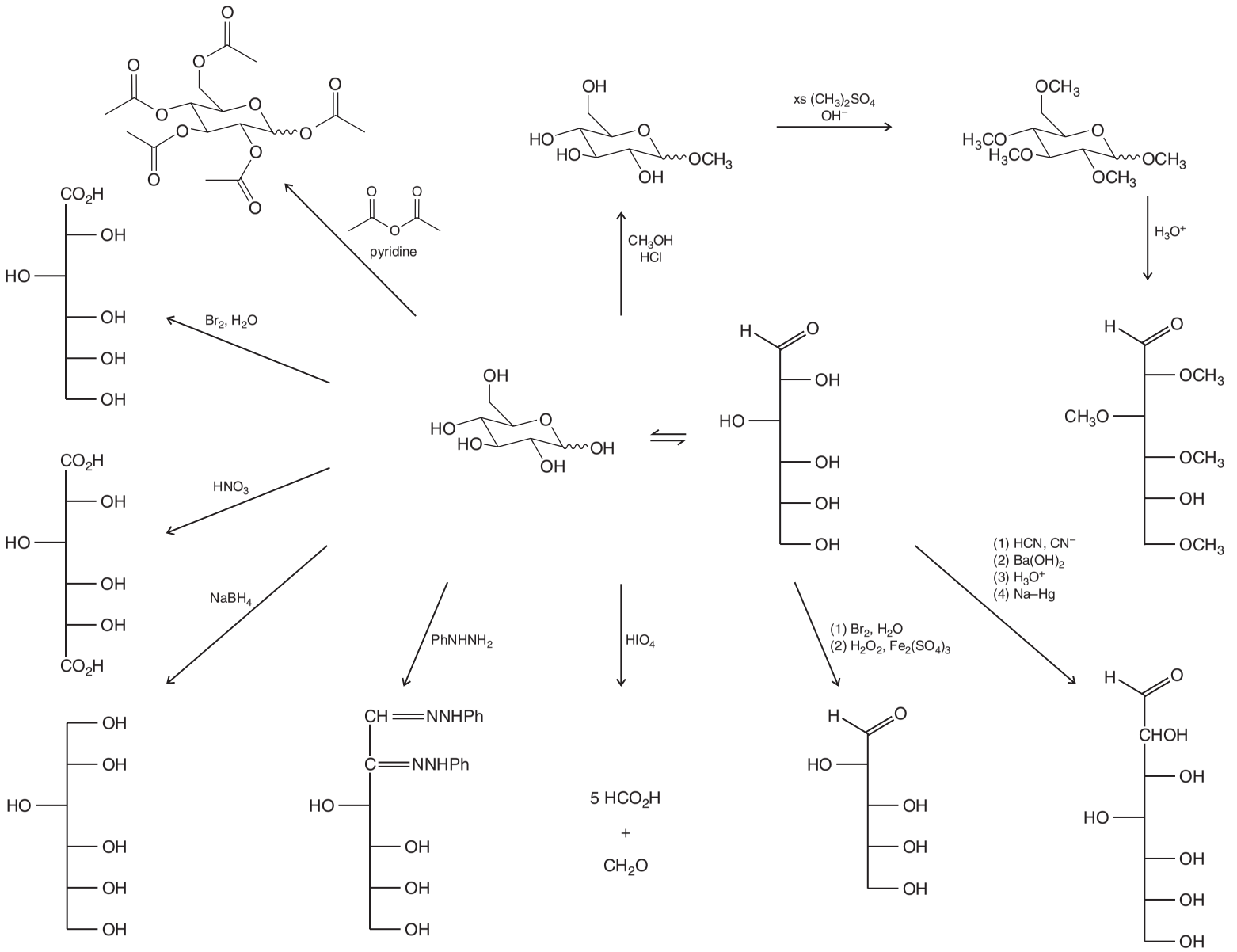
(d) Explain the transformation from **7** to **8**. Write out the structure of the compound in equilibrium with **7** that would be required for the process from **7** to **8** to occur. What is the name given to the reaction from this intermediate to **8**?

(e) Label the carbon atoms of D-mannose and **1** by number or letter so as to show which atoms in **1** came from which atoms of D-mannose.

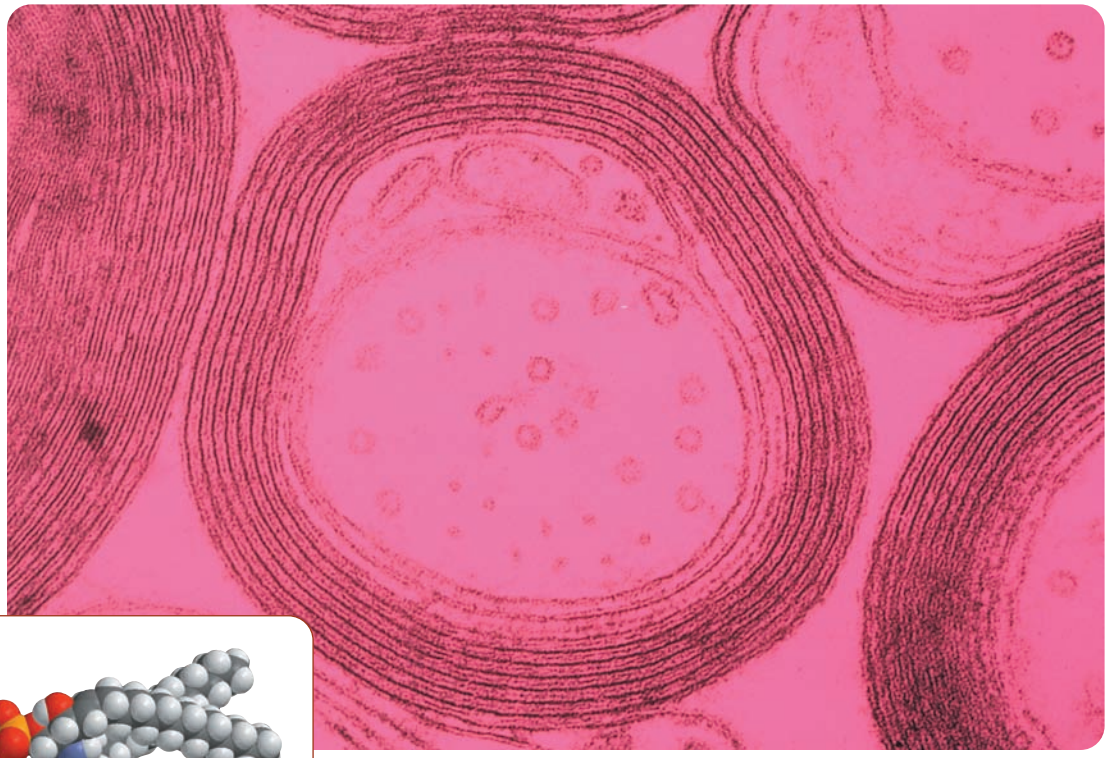


Scheme 1 The synthesis of D-(−)-[1,7-^{13}C]-shikimic acid (1) by H. G. Floss, based on the route of Fleet et al. Conditions: (a) acetone, HA; (b) BnCl, NaH; (c) HCl, aq. MeOH; (d) $NaIO_4$; (e) $NaBH_4$; (f) $(CF_3SO_2)_2O$, pyridine; (g) **9**, NaH; (h) $HCOO^-NH_4^+$, Pd/C; (i) NaH; (j) 60% aq. CF_3COOH.

A Summary of Reactions Involving Monosaccharides

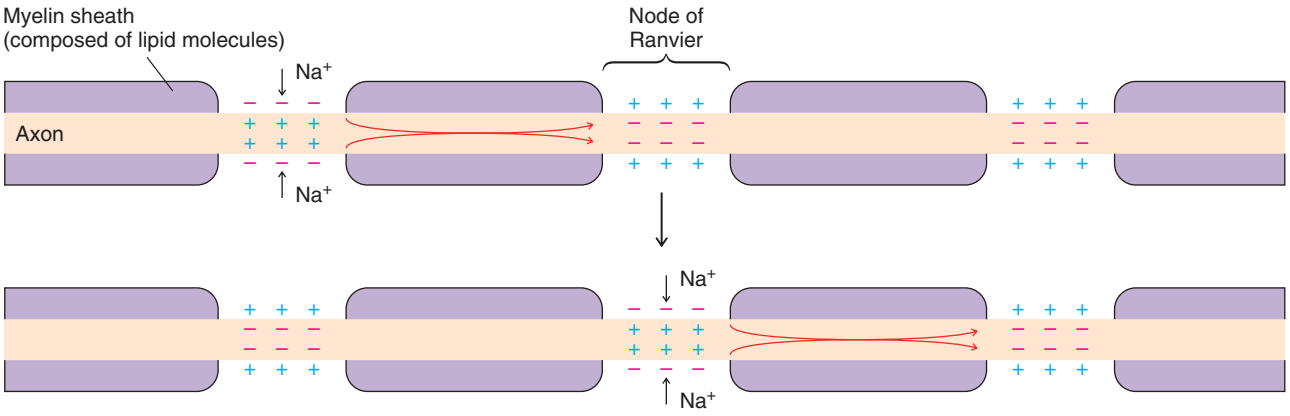


23 Lipids



A sphingomyelin molecule, found in myelin sheath membranes.

A bare wire conducting electricity will form a short circuit if it touches another conductor. This, of course, is why electrical wires are insulated. The axons of large neurons, the electrical conduits of the nervous system, are also insulated. Just as in electrical wires whose covering is an insulating sheath of plastic, a feature called the myelin sheath insulates the axons of many nerve cells from their extracellular environment. The myelin sheath is formed by the membrane of specialized cells, called Schwann cells, which grow around the axon and encircle it many times. In the structure of this membrane are molecules called lipids, a major component of which in myelin is sphingomyelin. A molecular model of sphingomyelin is shown above, and its structure is given in Section 23.6B.



(Reprinted with permission of John Wiley & Sons, Inc., from Voet, D. and Voet, J. G., *Biochemistry*, Second Edition. © 1995 Voet, D. and Voet, J. G.)

Wrapping of the axon by the Schwann cell membrane provides layer on layer of insulation by sphingomyelin and related lipid molecules. This is the key to the insulating property of the myelin sheath.

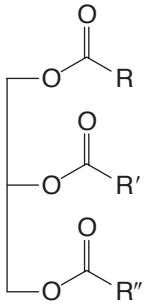
Unlike electrical wires that require insulation from end to end, the lipid layers of the myelin sheath are not a continuous insulator for the axon. Periodic gaps in the myelin sheath create nodes (called nodes of Ranvier) between which electrical signals of the nerve impulses hop along the axon. Propagation of nerve impulses in this way occurs at velocities up to 100 m s^{-1}, much faster than propagation in unmyelinated nerve fibers where this hopping effect is not possible. Impulse propagation in unmyelinated nerves is roughly 10 times slower than in myelinated nerves. The hopping of a nerve impulse between nodes is shown schematically in the diagram on the preceding page.

As you might expect, myelination of nerve fibers is crucial for proper neurological function. Multiple sclerosis, for example, is an autoimmune disease that causes demyelination of nerve cells, usually with very serious neurological consequences. Other conditions called sphingolipid storage diseases cause a buildup of various sphingolipids, which has various consequences. Examples of sphingolipid storage diseases are Tay-Sachs disease and Krabbe's disease. Both of these are fatal to children under the age of 3.

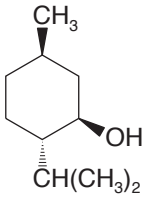
We shall see in this chapter that lipids come in a broad variety of classes—the sphingolipids mentioned here are but one example. We shall also see that the biological roles of lipids are even more varied and equally as fascinating as their structures.

23.1 Introduction

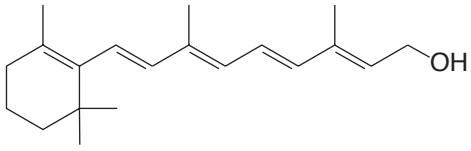
Lipids are compounds of biological origin that dissolve in nonpolar solvents, such as chloroform and diethyl ether. The name lipid comes from the Greek word *lipos*, for fat. Unlike carbohydrates and proteins, which are defined in terms of their structures, lipids are defined by the physical operation that we use to isolate them. Not surprisingly, then, lipids include a variety of structural types. Examples are the following:



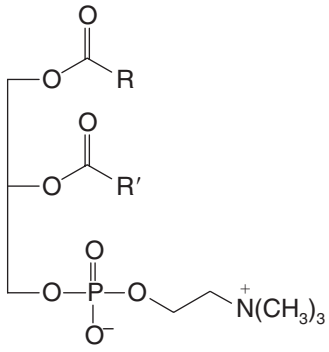
A fat or oil (a triacylglycerol)



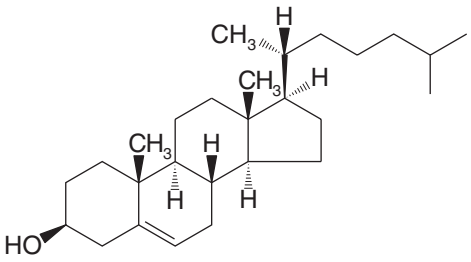
Menthol (a terpenoid)



Vitamin A (a terpenoid)

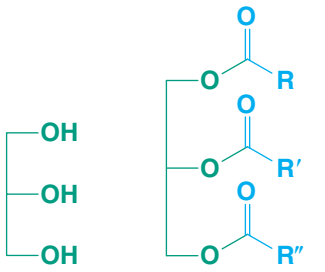


A lecithin (a phosphatide)



Cholesterol (a steroid)

23.2 Fatty Acids and Triacylglycerols



Glycerol (*a*) A triacylglycerol (*b*)

Figure 23.1 *(a)* Glycerol. *(b)* A triacylglycerol. The groups R, R′, and R″ are usually long-chain alkyl groups. R, R′, and R″ may also contain one or more carbon–carbon double bonds. In a triacylglycerol R, R′, and R″ may all be different.

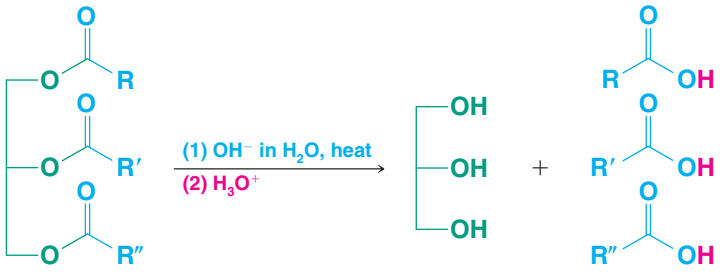
Only a small portion of the total lipid fraction obtained by extraction with a nonpolar solvent consists of long-chain carboxylic acids. Most of the carboxylic acids of biological origin are found as *esters of glycerol*, that is, as **triacylglycerols** (Fig. 23.1).*

Triacylglycerols are the oils of plants and the fats of animal origin. They include such common substances as peanut oil, soybean oil, corn oil, sunflower oil, butter, lard, and tallow.

- Triacylglycerols that are liquids at room temperature are generally called **oils;** those that are solids are called **fats**.

Triacylglycerols can be **simple triacylglycerols** in which all three acyl groups are the same. More commonly, however, the triacylglycerol is a **mixed triacylglycerol** in which the acyl groups are different.

- Hydrolysis of a fat or oil produces a mixture of **fatty acids**:



- Most natural fatty acids have **unbranched chains** and, because they are synthesized from two-carbon units, **they have an even number of carbon atoms**.

Table 23.1 lists some of the most common fatty acids, and Table 23.2 gives the fatty acid composition of a number of common fats and oils. Notice that in the **unsaturated fatty acids** in Table 23.1 **the double bonds are all cis.** Many naturally occurring fatty acids contain two or three double bonds. The fats or oils that these come from are called **polyunsaturated fats or oils.** The first double bond of an unsaturated fatty acid commonly occurs between C9 and C10; the remaining double bonds tend to begin with C12 and C15 (as in linoleic acid and linolenic acid). The double bonds, therefore, *are not conjugated*. Triple bonds rarely occur in fatty acids.

Helpful Hint

We saw how fatty acids are biosynthesized in two-carbon units in Special Topic E (*WileyPLUS*).

A saturated triacylglycerol

The carbon chains of **saturated fatty acids** can adopt many conformations but tend to be fully extended because this minimizes steric repulsions between neighboring methylene groups.

- Saturated fatty acids pack efficiently into crystals, and because dispersion force attractions are large, they have relatively high melting points. The melting points increase with increasing molecular weight.
- The cis configuration of the double bond of an unsaturated fatty acid puts a rigid bend in the carbon chain that interferes with crystal packing, causing reduced dispersion force attractions between molecules. Unsaturated fatty acids, consequently, have lower melting points.

Fatty acids known as omega-3 fatty acids are those where the third to last carbon in the chain is part of a carbon–carbon double bond. Long-chain omega-3 fatty acids incorporated in the diet are believed to have beneficial effects in terms of reducing the risk of fatal heart attack and easing certain autoimmune diseases, including rheumatoid arthritis and psoriasis.

*In the older literature triacylglycerols were referred to as triglycerides, or simply as glycerides. In IUPAC nomenclature, because they are esters of glycerol, they should be named as glyceryl trialkanoates, glyceryl trialkenoates, and so on.

TABLE 23.1 Common Fatty Acids

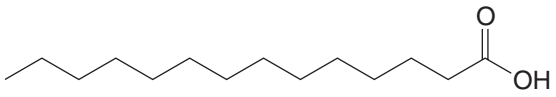
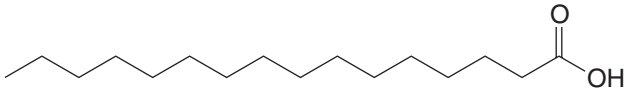
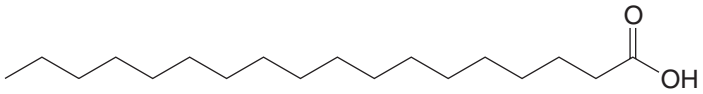
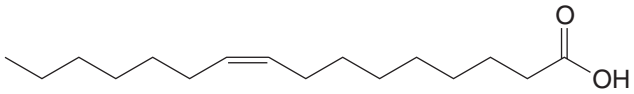
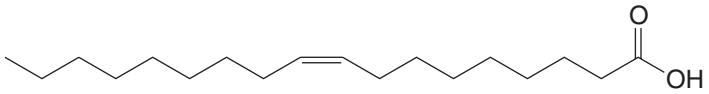
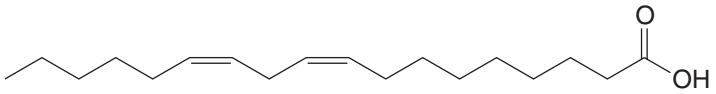
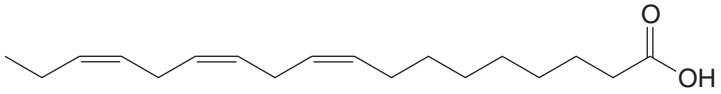
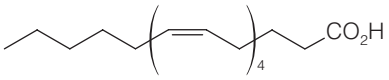
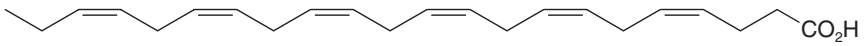
	mp (°C)
Saturated Carboxylic Acids	
 Myristic acid (tetradecanoic acid)	54
 Palmitic acid (hexadecanoic acid)	63
 Stearic acid (octadecanoic acid)	70
Unsaturated Carboxylic Acids	
 Palmitoleic acid (*cis*-9-hexadecenoic acid)	32
 Oleic acid (*cis*-9-octadecenoic acid)	4
 Linoleic acid (*cis,cis*-9,12-octadecadienoic acid)	−5
 Linoleic acid (*cis,cis,cis*-9,12,15-octadecatrienoic acid)	−11
 DHA, an omega-3 fatty acid [(4*Z*,7*Z*,10*Z*,13*Z*,16*Z*,19*Z*)-4,7,10,13,16,19-docosahexaenoic acid]	−44
 Arachidonic acid, an omega-6 fatty acid [(5*Z*,8*Z*,11*Z*,14*Z*)-5,8,11,14-eicosatetraenoic acid]	−49

TABLE 23.2 Fatty Acid Composition Obtained by Hydrolysis of Common Fats and Oils

	Average Composition of Fatty Acids (mol %)											
	Saturated								Unsaturated			
Fat or Oil	C_4 Butyric Acid	C_6 Caproic Acid	C_8 Caprylic Acid	C_{10} Capric Acid	C_{12} Lauric Acid	C_{14} Myristic Acid	C_{16} Palmitic Acid	C_{18} Stearic Acid	C_{16} Palmitoleic Acid	C_{18} Oleic Acid	C_{18} Linoleic Acid	C_{18} Linolenic Acid
Animal Fats												
Butter	3–4	1–2	0–1	2–3	2–5	8–15	25–29	9–12	4–6	18–33	2–4	
Lard						1–2	25–30	12–18	4–6	48–60	6–12	0–1
Beef tallow						2–5	24–34	15–30		35–45	1–3	0–1
Vegetable Oils												
Olive						0–1	5–15	1–4		67–84	8–12	
Peanut							7–12	2–6		30–60	20–38	
Corn						1–2	7–11	3–4	1–2	25–35	50–60	
Cottonseed						1–2	18–25	1–2	1–3	17–38	45–55	
Soybean						1–2	6–10	2–4		20–30	50–58	5–10
Linseed							4–7	2–4		14–30	14–25	45–60
Coconut		0–1	5–7	7–9	40–50	15–20	9–12	2–4	0–1	6–9	0–1	
Marine Oils												
Cod liver						5–7	8–10	0–1	18–22	27–33	27–32	

Oil from fish such as tuna and salmon is a good source of omega-3 fatty acids, including the C_{22} omega-3 fatty acid docosahexaenoic acid [DHA, whose full IUPAC name is (4*Z*,7*Z*,10*Z*,13*Z*,16*Z*,19*Z*)-4,7,10,13,16,19-docosahexaenoic acid]. DHA is also found in breast milk, gray matter of the brain, and retinal tissue.



(4*Z*,7*Z*,10*Z*,13*Z*,16*Z*,19*Z*)-4,7,10,13,16,19-Docosahexaenoic acid (DHA, an omega-3 fatty acid)

What we have just said about the fatty acids applies to the triacylglycerols as well. Triacylglycerols made up of largely saturated fatty acids have high melting points and are solids at room temperature. They are what we call *fats*. Triacylglycerols with a high proportion of unsaturated and polyunsaturated fatty acids have lower melting points. They are *oils*. Figure 23.2 shows how the introduction of a single cis double bond affects the shape of a triacylglycerol and how catalytic hydrogenation can be used to convert an unsaturated triacylglycerol into a saturated one.

23.2A Hydrogenation of Triacylglycerols

Solid commercial cooking fats are manufactured by partial hydrogenation of vegetable oils. The result is the familiar "partially hydrogenated fat" present in so many prepared foods. Complete hydrogenation of the oil is avoided because a completely saturated triacylglycerol is very hard and brittle. Typically, the vegetable oil is hydrogenated until a semisolid of appealing consistency is obtained. One commercial advantage of partial hydrogenation is to give the fat a longer shelf life. Polyunsaturated oils tend to react by autoxidation (Section 10.11D), causing them to become rancid. One problem with partial hydrogenation, however, is that the catalyst isomerizes some of the unreacted double bonds from the natural cis arrangement to the unnatural trans arrangement, and there is accumulating evidence that trans fats are associated with an increased risk of cardiovascular disease.

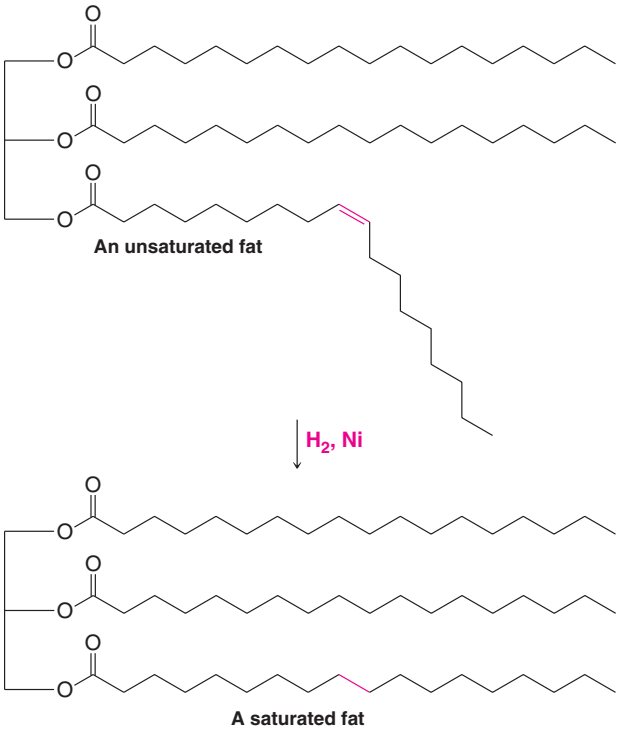


Figure 23.2 Two typical triacylglycerols, one unsaturated and one saturated. The cis double bond of the unsaturated triacylglycerol interferes with efficient crystal packing and causes an unsaturated fat to have a lower melting point. Hydrogenation of the double bond causes an unsaturated triacylglycerol to become saturated.

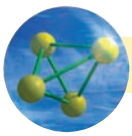
23.2B Biological Functions of Triacylglycerols

The primary function of triacylglycerols in animals is as an energy reserve. When triacylglycerols are converted to carbon dioxide and water by biochemical reactions (i.e., when triacylglycerols are *metabolized*), they yield more than twice as many kilocalories per gram as do carbohydrates or proteins. This is largely because of the high proportion of carbon–hydrogen bonds per molecule.

In animals, specialized cells called **adipocytes** (fat cells) synthesize and store triacylglycerols. The tissue containing these cells, adipose tissue, is most abundant in the abdominal cavity and in the subcutaneous layer. Men have a fat content of about 21%, women about 26%. This fat content is sufficient to enable us to survive starvation for 2–3 months. By contrast, glycogen, our carbohydrate reserve, can provide only one day's energy need.

All of the saturated triacylglycerols of the body, and some of the unsaturated ones, can be synthesized from carbohydrates and proteins. Certain polyunsaturated fatty acids, however, are essential in the diets of higher animals.

The amount of fat in the diet, especially the proportion of saturated fat, has been a health concern for many years. There is compelling evidence that too much saturated fat in the diet is a factor in the development of heart disease and cancer.



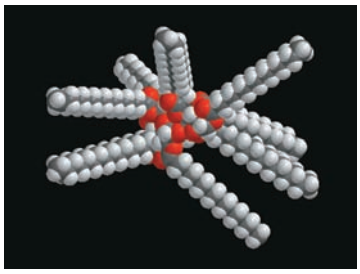
THE CHEMISTRY OF . . .

Olestra and Other Fat Substitutes

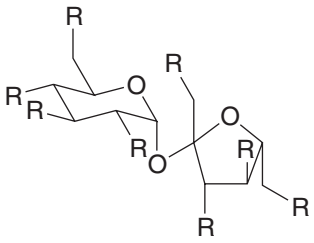
Olestra is a zero-calorie commercial fat substitute with the look and feel of natural fats. It is a synthetic compound whose structure involves a novel combination of natural components. The core of olestra is derived from sucrose, ordinary table sugar. Six to eight of the hydroxyl groups on the sucrose framework have long-chain carboxylic acids (fatty acids) appended to them by ester linkages. These fatty acids are from C_8 to C_{22} in length. In the industrial synthesis of olestra, these fatty acids derive from cottonseed or soybean oil.

(Structure at right used with permission from the Journal of Chemical Education, Vol. 74, No. 4, 1997, pp. 370–372; copyright © 1997, Division of Chemical Education, Inc.)

A food product made with olestra.



Olestra.



Olestra
Six to eight of the R groups are fatty acid esters, the remainder being hydroxyl groups.

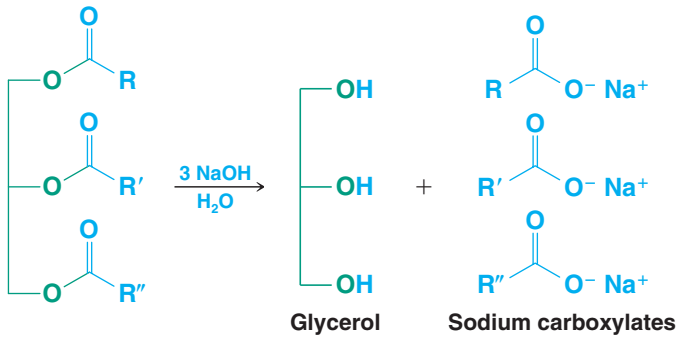
The presence of fatty acid esters in olestra bestows on it the taste and culinary properties of an ordinary fat. Yet, olestra is not digestible like a typical fat. This is because the steric bulk of olestra renders it unacceptable to the enzymes that catalyze hydrolysis of ordinary fats. Olestra passes through the digestive tract unchanged and thereby adds no calories to the diet. As it does so, however, olestra associates with and carries away some of the lipid-soluble vitamins, namely, vitamins A, D, E, and K. Foods prepared with olestra are supplemented with these vitamins to compensate for any loss that may result from their extraction by olestra. Studies conducted since olestra's approval have demonstrated that people report no more bothersome digestive effects when eating Olean (the trademark name for olestra) snacks than they do when eating full-fat chips.

Many other fat substitutes have received consideration. Among these are polyglycerol esters, which presumably by their steric bulk would also be undigestible, like the polyester olestra. Another approach to low-calorie fats, already in commercial use, involves replacement of some long-chain carboxylic acids on the glycerol backbone with medium- or short-chain carboxylic acids (C_2 to C_4). These compounds provide fewer calories because each CH_2 group that is absent from the glycerol ester (as compared to long-chain fatty acids) reduces the amount of energy (calories) liberated when that compound is metabolized. The calorie content of a given glycerol ester can essentially be tailored to provide a desired calorie output, simply by adjusting the ratio of long-chain to medium- and short-chain carboxylic acids. Still other low-calorie fat substitutes are carbohydrate- and protein-based compounds. These materials act by generating a similar gustatory response to that of fat, but for various reasons produce fewer calories.

Structure of olestra adapted by permission from *Journal of Chemical Education*, Vol. 74, No. 4, 1997, pp. 370–372. Copyright 1997, Division of Chemical Education, Inc.

23.2C Saponification of Triacylglycerols

- **Saponification** is the alkaline hydrolysis of triacylglycerols, leading to glycerol and a mixture of salts of long-chain carboxylic acids:



Glycerol **Sodium carboxylates "soap"**

These salts of long-chain carboxylic acids are **soaps,** and this saponification reaction is the way most soaps are manufactured. Fats and oils are boiled in aqueous sodium hydroxide until hydrolysis is complete. Adding sodium chloride to the mixture then causes the soap to precipitate. (After the soap has been separated, glycerol can be isolated from the aqueous phase by distillation.) Crude soaps are usually purified by several reprecipitations. Perfumes can be added if a toilet soap is the desired product. Sand, sodium carbonate, and

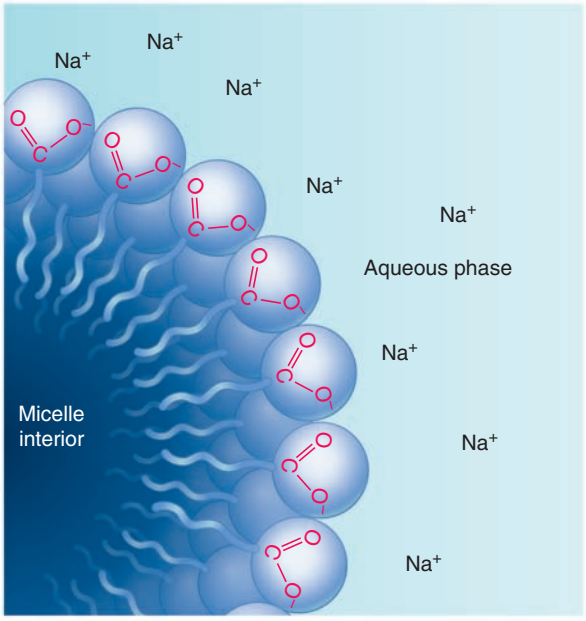


Figure 23.3 A portion of a soap micelle showing its interface with the polar dispersing medium. (Reprinted with permission of John Wiley & Sons, Inc., from Karp, G., *Cell and Molecular Biology: Concepts and Experiments*, Fourth Edition, Copyright 1999.)

other fillers can be added to make a scouring soap, and air can be blown into the molten soap if the manufacturer wants to market a soap that floats.

The sodium salts of long-chain carboxylic acids (soaps) are almost completely miscible with water. However, they do not dissolve as we might expect, that is, as individual ions. Except in very dilute solutions, soaps exists as **micelles** (Fig. 23.3). Soap micelles are usually spherical clusters of carboxylate anions that are dispersed throughout the aqueous phase. The carboxylate anions are packed together with their negatively charged (and thus, *polar*) carboxylate groups at the surface and with their nonpolar hydrocarbon chains on the interior. The sodium ions are scattered throughout the aqueous phase as individual solvated ions.

Micelle formation accounts for the fact that soaps dissolve in water. The nonpolar (and thus **hydrophobic**) alkyl chains of the soap remain in a nonpolar environment—in the interior of the micelle. The polar (and therefore **hydrophilic**) carboxylate groups are exposed to a polar environment—that of the aqueous phase. Because the surfaces of the micelles are negatively charged, individual micelles repel each other and remain dispersed throughout the aqueous phase.

Soaps serve their function as "dirt removers" in a similar way. Most dirt particles (e.g., on the skin) become surrounded by a layer of an oil or fat. Water molecules alone are unable to disperse these greasy globules because they are unable to penetrate the oily layer and separate the individual particles from each other or from the surface to which they are stuck. Soap solutions, however, *are* able to separate the individual particles because their hydrocarbon chains can "dissolve" in the oily layer (Fig. 23.4). As this happens, each individual particle develops an outer layer of carboxylate anions and presents the aqueous phase with a much more compatible exterior—a polar surface. The individual globules now repel each other and thus become dispersed throughout the aqueous phase. Shortly thereafter, they make their way down the drain.

Synthetic detergents (Fig. 23.5) function in the same way as soaps; they have long nonpolar alkane chains with polar groups at the end. The polar groups of most synthetic detergents are sodium sulfonates or sodium sulfates. (At one time, extensive use was made of synthetic detergents with highly branched alkyl groups. These detergents proved to be nonbiodegradable, and their use was discontinued.)

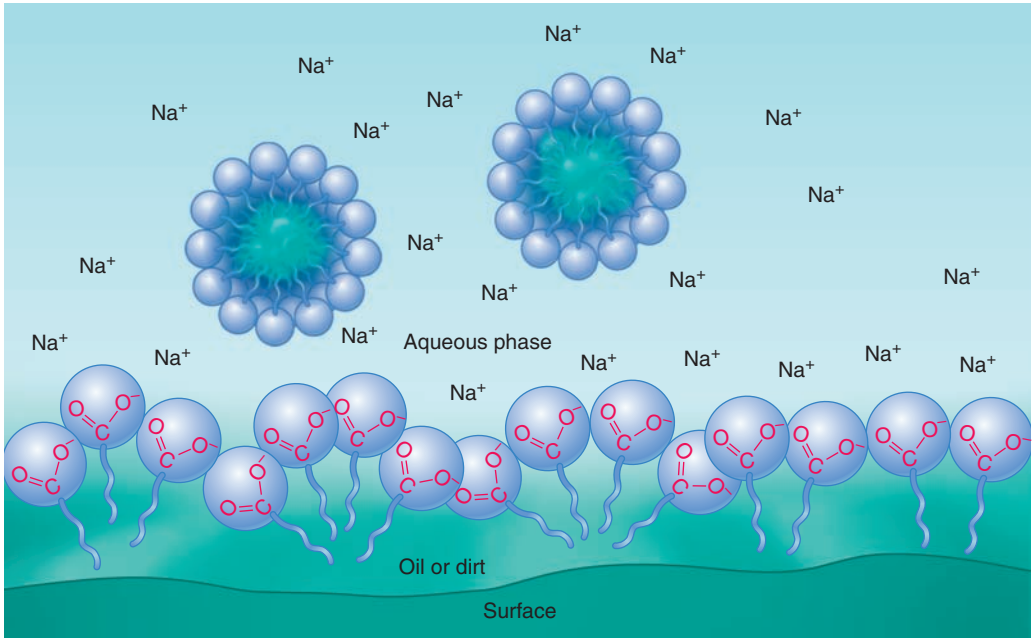


Figure 23.4 Dispersal of a hydrophobic material (e.g., oil, grease, or fat) by a soap. (Adapted with permission of John Wiley & Sons, Inc., from Karp, G., *Cell and Molecular Biology: Concepts and Experiments*, Fourth Edition, Copyright 1999.)

Synthetic detergents offer an advantage over soaps; they function well in "hard" water, that is, water containing Ca^{2+}, Fe^{2+}, Fe^{3+}, and Mg^{2+} ions. Calcium, iron, and magnesium salts of alkanesulfonates and alkyl hydrogen sulfates are largely water soluble, and thus synthetic detergents remain in solution. Soaps, by contrast, form precipitates—the ring around the bathtub—when they are used in hard water.

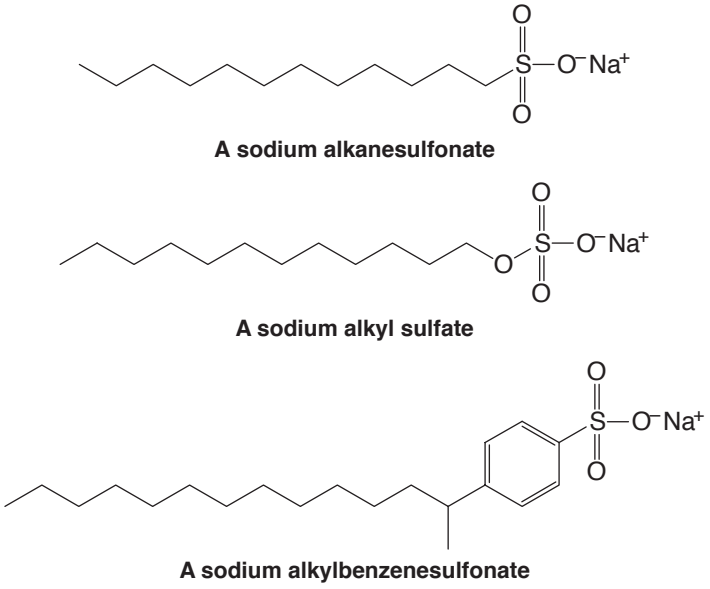
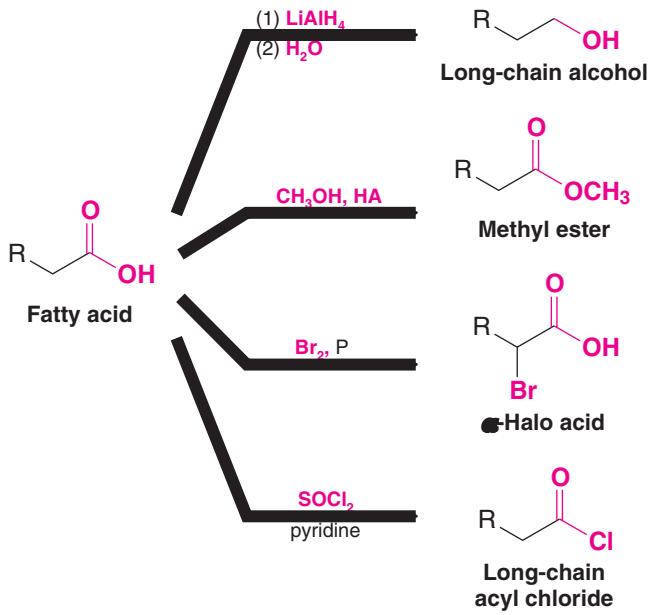


Figure 23.5 Typical synthetic detergents.

23.2D Reactions of the Carboxyl Group of Fatty Acids

Fatty acids, as we might expect, undergo reactions typical of carboxylic acids (see Chapter 17). They react with $LiAlH_4$ to form alcohols, with alcohols and mineral acid to form esters, with bromine and phosphorus to form α-halo acids, and with thionyl chloride to form acyl chlorides:

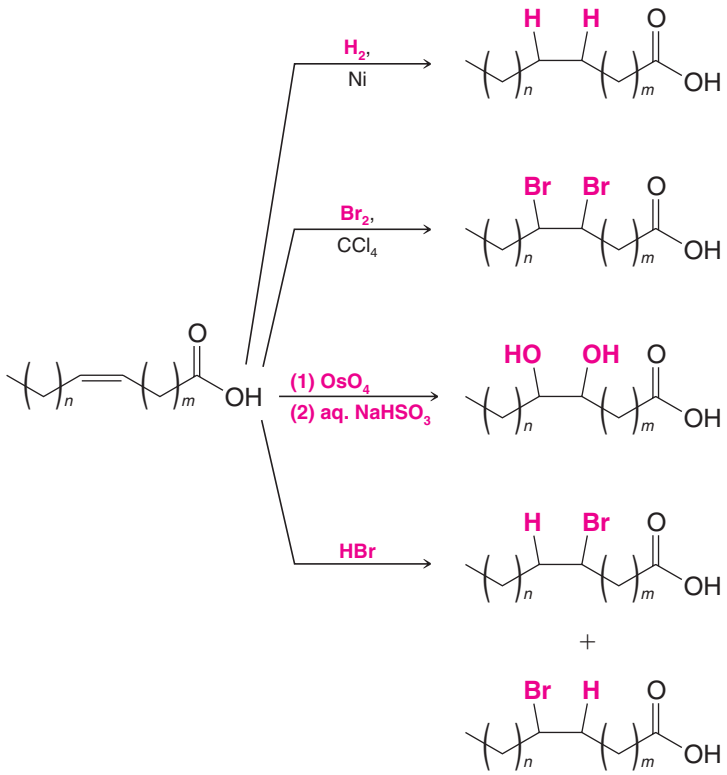


Helpful Hint

The reactions presented in Sections 23.2D and 23.2E in the context of fatty acids are the same as those we studied in earlier chapters regarding carboxylic acids and alkenes.

23.2E Reactions of the Alkenyl Chain of Unsaturated Fatty Acids

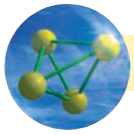
The double bonds of the carbon chains of fatty acids undergo characteristic alkene addition reactions (see Chapters 7 and 8):



Review Problem 23.1

(a) How many stereoisomers are possible for 9,10-dibromohexadecanoic acid?

(b) The addition of bromine to palmitoleic acid yields primarily one set of enantiomers, (±)-*threo*-9,10-dibromohexadecanoic acid. The addition of bromine is an anti addition to the double bond (i.e., it apparently takes place through a bromonium ion intermediate). Taking into account the cis stereochemistry of the double bond of palmitoleic acid and the stereochemistry of the bromine addition, write three-dimensional structures for the (±)-*threo*-9,10-dibromohexadecanoic acids.

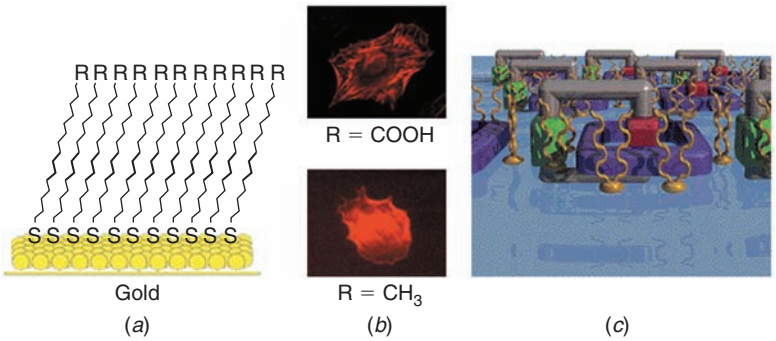


THE CHEMISTRY OF . . .

Self-Assembled Monolayers—Lipids in Materials Science and Bioengineering

The graphic shown below (a) depicts a self-assembled monolayer of alkanethiol molecules on a gold surface. The alkanethiol molecules spontaneously form a layer that is one molecule thick (a monolayer) because they are tethered to the gold surface at one end by a covalent bond to the metal and because van der Waals intermolecular forces between the long alkane chains cause them to align next to each other in an approximately perpendicular orientation to the gold surface. Many researchers are exploiting self-assembled monolayers (SAMs) for the preparation of surfaces that have specific uses in medicine, computing, and telecommunications. One example in biomedical engineering that may lead to advances in surgery involves testing cells for their response to SAMs with varying head groups. By varying the structure of the exposed head group of the monolayer, it may be possible to create materials that have either affinity for or resistance against cell binding (b). Such properties could be useful in organ transplants for inhibiting rejection by cells of the immune system or in prosthesis surgeries where the binding of tissue to the artificial device is desired.

Monolayers called Langmuir–Blodgett (LB) films also involve self-assembly of molecules on a surface. In this case, however, the molecules do not become covalently attached to the surface. These LB films are inherently less stable than covalently bonded monolayers, but they have characteristics that are useful for certain applications in nanotechnology. For example, an LB film made from phospholipid (Section 23.6) and catenane molecules was used in making the array of molecular switches we discussed in "The Chemistry of . . . Nanoscale Motors and Molecular Switches" (Chapter 4). This LB monolayer (c) was formed at a water–air interface where the polar phosphate head groups of the phospholipid buried themselves in water and the hydrophobic carbon tails projected out into the air. Interspersed among them were the catenane molecules. In later steps, this monolayer was lifted from the water–air surface and transferred onto a solid gold surface.



(*a*) A self-assembled monolayer of alkanethiol molecules on a gold surface (R = CH_3 or COOH). (*b*) Spreading of a Swiss 3T3 fibroblast cell plated on a COOH-terminated self-assembled monolayer (top) indicates effective signaling on the surface. The fibroblast cell on a CH_3-terminated monolayer (bottom) curls away from surface. The cells were stained with a rhodamine-tagged toxin that binds to filamentous actin and then were imaged under fluorescent light. (*c*) A Langmuir–Blodgett (LB) film formed from phospholipid molecules (golden color) and catenane molecules (purple and gray with green and red groups) at an air–water interface. (*Image of switching devices based on interlocked molecules* reprinted with permission from Pease, A.R., Jeppensen, J.E., et al., *Accounts of Chemical Research, Vol. 34*, No. 6, p. 433, Figure 8C, June 2001. Copyright 2001 American Chemical Society.)

23.3 Terpenes and Terpenoids

People have isolated organic compounds from plants since antiquity. By gently heating or by steam distilling certain plant materials, one can obtain mixtures of odoriferous compounds known as **essential oils**. These compounds have had a variety of uses, particularly in early medicine and in the making of perfumes.

As the science of organic chemistry developed, chemists separated the various components of these mixtures and determined their molecular formulas and, later, their structural formulas. Even today these natural products offer challenging problems for chemists interested in structure determination and synthesis. Research in this area has also given us important information about the ways the plants themselves synthesize these compounds.

Helpful Hint

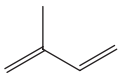
Terpene biosynthesis was described in Special Topic E (*WileyPLUS*)

- Hydrocarbons known generally as **terpenes** and oxygen-containing compounds called **terpenoids** are the most important constituents of essential oils.
- Most terpenes have skeletons of 10, 15, 20, or 30 carbon atoms and are classified in the following way:

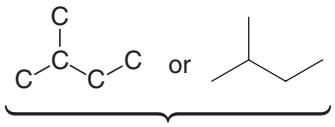
Number of Carbon Atoms	Class
10	Monoterpenes
15	Sesquiterpenes
20	Diterpenes
30	Triterpenes

- One can view terpenes as being built up from two or more C_5 units known as **isoprene units**. Isoprene is 2-methyl-1,3-butadiene.

Isoprene and the isoprene unit can be represented in various ways:

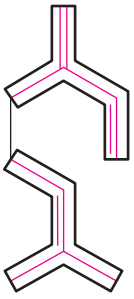


2-Methyl-1,3-butadiene (isoprene)

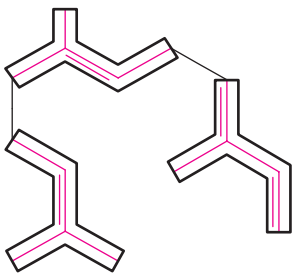


An isoprene unit

We now know that plants do not synthesize terpenes from isoprene (see Special Topic E, *WileyPLUS*). However, recognition of the isoprene unit as a component of the structure of terpenes has been a great aid in elucidating their structures. We can see how if we examine the following structures:



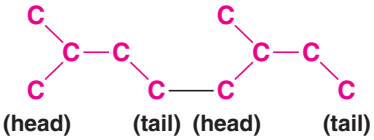
Myrcene (isolated from bay oil)



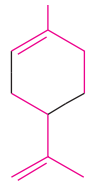
β-Farnesene (from natural coating of apples)

The isoprene units in myrcene and β-Farnesene are outlined.

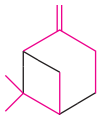
By the outlines in the formulas above, we can see that the monoterpene (myrcene) has two isoprene units; the sesquiterpene (α-farnesene) has three. In both compounds the isoprene units are linked head to tail:



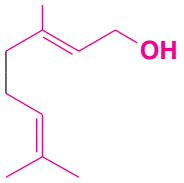
Many terpenes also have isoprene units linked in rings, and others (terpenoids) contain oxygen:



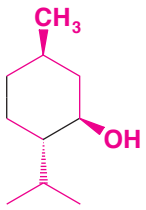
Limonene
(from oil of lemon or orange)



β-Pinene
(from oil of turpentine)



Geraniol
(from roses and other flowers)



Menthol
(from peppermint)

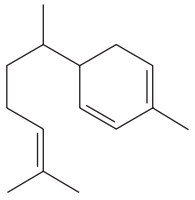
Solved Problem 23.1

Hydrogenation of the sesquiterpene caryophyllene produces a compound with the molecular formula $C_{15}H_{28}$. What information does this provide about the structure of caryophyllene?

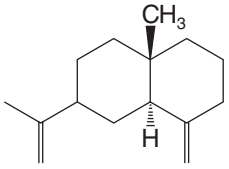
STRATEGY AND ANSWER Caryophyllene has the molecular formula $C_{15}H_{24}$, and therefore an index of hydrogen deficiency (IHD) of 4. Its reaction with two molar equivalents of hydrogen suggests that caryophyllene has two double bonds or one triple bond, accounting for two of the four units of hydrogen deficiency. The remaining two units of hydrogen deficiency are due to rings. (The structure of caryophyllene is given in Review Problem 23.2.)

Review Problem 23.2

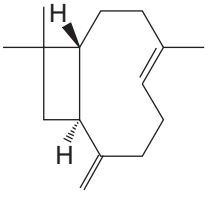
(a) Show the isoprene units in each of the following terpenes. **(b)** Classify each as a monoterpene, sesquiterpene, diterpene, and so on.



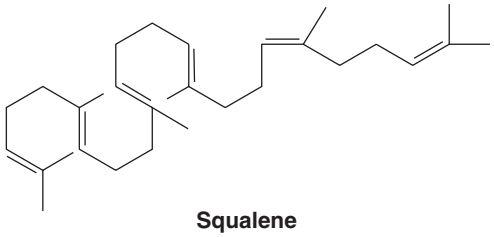
Zingiberene
(from oil of ginger)



β-Selinene
(from oil of celery)



Caryophyllene
(from oil of cloves)

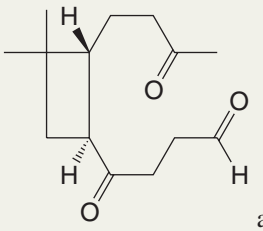


Squalene
(from shark liver oil)

Solved Problem 23.2

What products would you expect to obtain if caryophyllene were subjected to ozonolysis followed by workup with dimethyl sulfide?

ANSWER



and HCHO (formaldehyde)

Review Problem 23.3

What products would you expect to obtain if each of the following terpenes were subjected to ozonolysis and subsequent treatment with dimethyl sulfide?

(a) Myrcene **(b)** Limonene **(c)** α-Farnesene **(d)** Geraniol **(e)** Squalene

Review Problem 23.4

Give structural formulas for the products that you would expect from the following reactions:

(a) β-Pinene $\xrightarrow{KMnO_4,\ heat}$

(b) Zingiberene $\xrightarrow{H_2,\ Pt}$

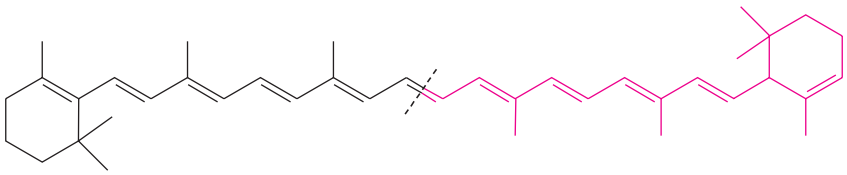
(c) Caryophyllene $\xrightarrow{HCl}$

(d) β-Selinene $\xrightarrow[(2)\ H_2O_2,\ OH^-]{(1)\ BH_3:THF\ (2\ eqviv.)}$

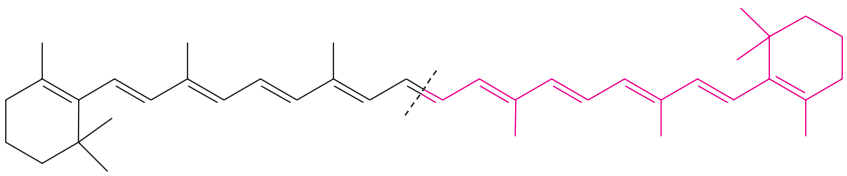
Review Problem 23.5

What simple chemical test could you use to distinguish between geraniol and menthol?

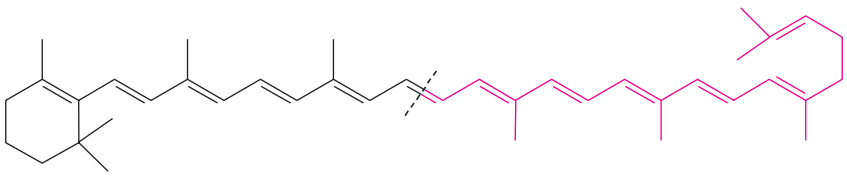
The carotenes are tetraterpenes. They can be thought of as two diterpenes linked in tail-to-tail fashion:



α-Carotene

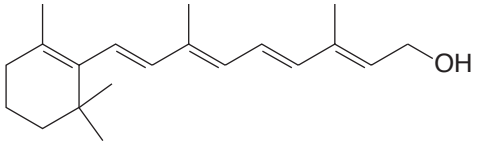


β-Carotene



γ-Carotene

The carotenes are present in almost all green plants. In animals, all three carotenes serve as precursors for vitamin A, for they all can be converted to vitamin A by enzymes in the liver.

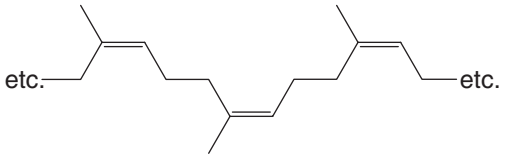


Vitamin A

In this conversion, one molecule of β-carotene yields two of vitamin A; α- and γ-carotene give only one. Vitamin A is important not only in vision but in many other ways as well. For example, young animals whose diets are deficient in vitamin A fail to grow. Vitamin A, β-carotene, and vitamin E ("The Chemistry of . . . Antioxidants," Section 10.11) are important lipid-soluble antioxidants, as well.

23.3A Natural Rubber

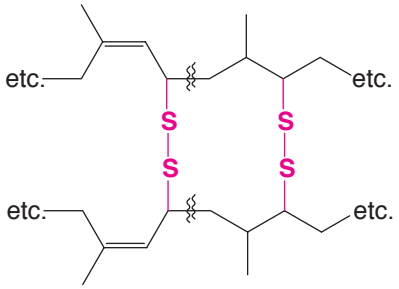
Natural rubber can be viewed as a 1,4-addition polymer of isoprene. In fact, pyrolysis degrades natural rubber to isoprene. Pyrolysis (Greek: *pyros*, a fire, + *lysis*) is the heating of a substance in the absence of air until it decomposes. The isoprene units of natural rubber are all linked in a head-to-tail fashion, and all of the double bonds are cis:



Natural rubber
(*cis*-1,4-polyisoprene)

Ziegler–Natta catalysts (see Special Topic B) make it possible to polymerize isoprene and obtain a synthetic product that is identical with the rubber obtained from natural sources.

Pure natural rubber is soft and tacky. To be useful, natural rubber has to be *vulcanized*. In vulcanization, natural rubber is heated with sulfur. A reaction takes place that produces cross-links between the *cis*-polyisoprene chains and makes the rubber much harder. Sulfur reacts both at the double bonds and at allylic hydrogen atoms:



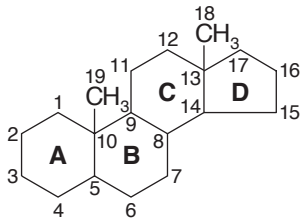
Vulcanized rubber

23.4 Steroids

The lipid fractions obtained from plants and animals contain another important group of compounds known as **steroids**. Steroids are important "biological regulators" that nearly always show dramatic physiological effects when they are administered to living organisms. Among these important compounds are male and female sex hormones, adrenocortical hormones, D vitamins, the bile acids, and certain cardiac poisons.

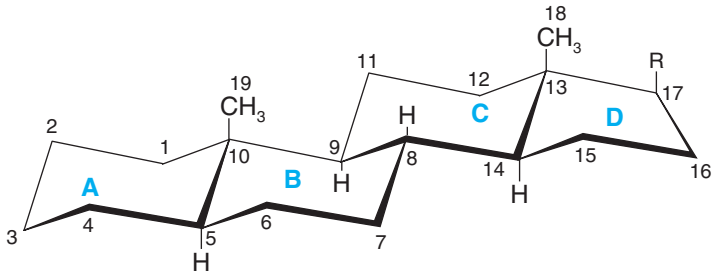
23.4A Structure and Systematic Nomenclature of Steroids

Steroids are derivatives of the following perhydrocyclopentanophenanthrene ring system:

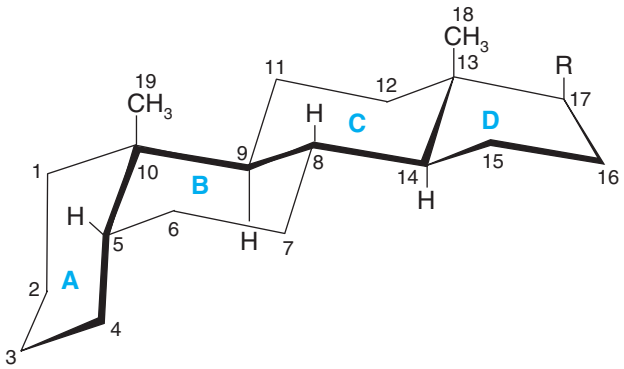


The carbon atoms of this ring system are numbered as shown. The four rings are designated with letters. In most steroids the B,C and C,D ring junctions are trans. The A,B ring junction, however, may be either cis or trans, and this possibility gives rise to two general groups of steroids having the three-dimensional structures shown in Fig. 23.6.

The methyl groups that are attached at points of ring junction (i.e., those numbered 18 and 19) are called **angular methyl groups,** and they serve as important reference points for stereochemical designations. The angular methyl groups protrude above the general plane of the ring system when it is written in the manner shown in Fig. 23.6. By convention, other groups that lie on the same general side of the molecule as the angular methyl groups (i.e., on the top side) are designated ***β* substituents** (these are written with a solid wedge). Groups that lie generally on the bottom (i.e., are trans to the angular methyl groups) are designated ***α* substituents** (these are written with a dashed wedge). When α and β designations are applied to the hydrogen atom at position 5, the ring system in which the A,B ring junction is trans becomes the 5α series; the ring system in which the A,B ring junction is cis becomes the 5β series.



5α Series of steroids
(All ring junctions are trans.)



5β Series of steroids
(A,B ring junction is cis.)

Figure 23.6 The basic ring systems of the 5α and 5β series of steroids.

Helpful Hint

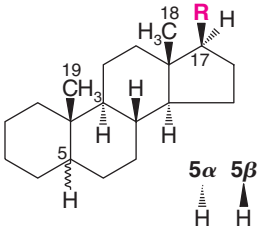
Build handheld molecular models of the 5α and 5β series of steroids and use them to explore the structures of steroids discussed in this chapter.

Review Problem 23.6

Draw the two basic ring systems given in Fig. 23.6 for the 5α and 5β series showing all hydrogen atoms of the cyclohexane rings. Label each hydrogen atom as to whether it is axial or equatorial.

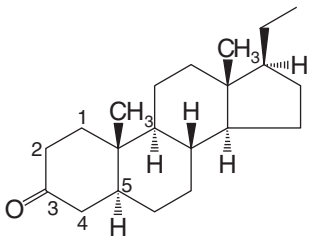
In systematic nomenclature the nature of the R group at position 17 determines (primarily) the base name of an individual steroid. These names are derived from the steroid hydrocarbon names given in Table 23.3.

TABLE 23.3 Names of Steroid Hydrocarbons

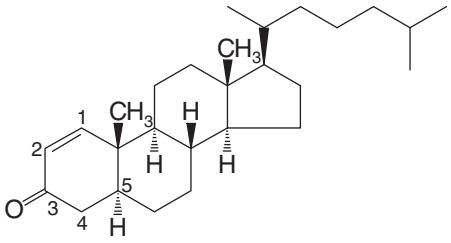


R	Name
H	Androstane
H (with —H also replacing $-\overset{19}{C}H_3$)	Estrane
20, 21	Pregnane
21, 20, 22, 23, 24	Cholane
21, 20, 22, 23, 24, 25, 26, 27	Cholestane

The following two examples illustrate the way these base names are used:



5α-Pregnan-3-one



5α-Cholest-1-en-3-one

We shall see that many steroids also have common names and that the names of the steroid hydrocarbons given in Table 23.3 are derived from these common names.

Review Problem 23.7

(a) Androsterone, a secondary male sex hormone, has the systematic name 3α-hydroxy-5α-androstan-17-one. Give a three-dimensional formula for androsterone.

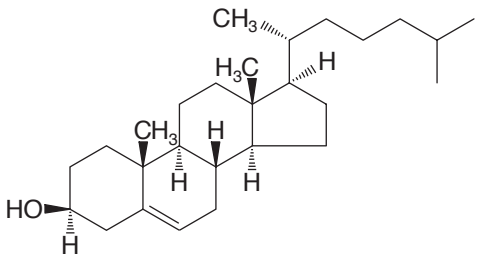
(b) Norethynodrel, a synthetic steroid that has been widely used in oral contraceptives, has the systematic name 17α-ethynyl-17β-hydroxy-5(10)-estren-3-one. Give a three-dimensional formula for norethynodrel.

23.4B Cholesterol

Cholesterol, one of the most widely occurring steroids, can be isolated by extraction of nearly all animal tissues. Human gallstones are a particularly rich source.

Cholesterol was first isolated in 1770. In the 1920s, two German chemists, Adolf Windaus (University of Göttingen) and Heinrich Wieland (University of Munich), were responsible for outlining a structure for cholesterol; they received Nobel prizes for their work in 1927 and 1928.*

Part of the difficulty in assigning an absolute structure to cholesterol is that cholesterol contains *eight* tetrahedral chirality centers. This feature means that 2^8, or 256, possible stereoisomeric forms of the basic structure are possible, *only one of which is cholesterol*:



5-Cholesten-3 β-ol
(absolute configuration of cholesterol)

Helpful Hint

We saw how cholesterol is biosynthesized in "The Chemistry of . . . Cholesterol Biosynthesis" in *WileyPLUS* materials *for* Chapter 8.

Designate with asterisks the eight chirality centers of cholesterol.

Review Problem 23.8

Cholesterol occurs widely in the human body, but not all of the biological functions of cholesterol are yet known. Cholesterol is known to serve as an intermediate in the biosynthesis of all of the steroids of the body. Cholesterol, therefore, is essential to life. We do not need to have cholesterol in our diet, however, because our body can synthesize all we need. When we ingest cholesterol, our body synthesizes less than if we ate none at all, but the total cholesterol is more than if we ate none at all. Far more cholesterol is present in the body than is necessary for steroid biosynthesis. High levels of blood cholesterol have been implicated in the development of arteriosclerosis (hardening of the arteries) and in heart attacks that occur when cholesterol-containing plaques block arteries of the heart. Considerable research is being carried out in the area of cholesterol metabolism with the hope of finding ways of minimizing cholesterol levels through the use of dietary adjustments or drugs.

It is important to note that, in common language, "cholesterol" does not necessarily refer only to the pure compound that chemists call cholesterol, but often refers instead to mixtures that contain cholesterol, other lipids, and proteins. These aggregates are called chylomicrons, high-density lipoproteins (HDLs), and low-density lipoproteins (LDLs). They have structures generally resembling globular micelles, and they are the vehicles by which cholesterol is transported through the aqueous environment of the body. Hydrophilic groups of their constituent proteins and phospholipids, and cholesterol hydroxyl substituents are oriented outward toward the water medium so as to facilitate transport of the lipids through the circulatory system. HDLs (the "good cholesterol") carry lipids from the tissues to the liver for degradation and excretion. LDL ("bad cholesterol") carries biosynthesized lipids from the liver to the tissues (see Fig. 23.7). Chylomicrons transport dietary lipids from the intestines to the tissues.

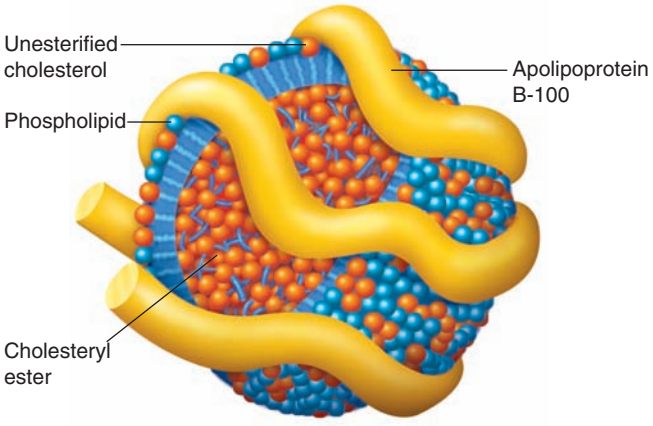
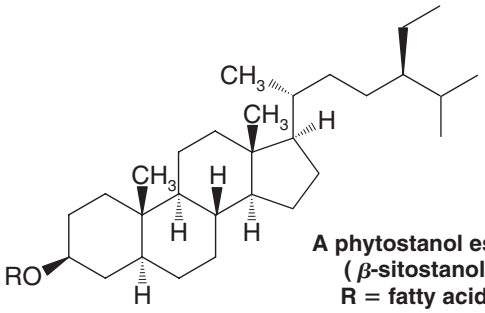


Figure 23.7 An LDL showing a core of cholesterol esters and a shell of phospholipids and unesterified cholesterol (hydroxyl groups exposed), wrapped in an apolipoprotein. The phospholipid head groups and hydrophilic residues of the protein support the water compatibility of the LDL particle. (Reprinted with permission of John Wiley & Sons, Inc., from Voet, D. and Voet, J. G., *Biochemistry*, Second Edition. © 1995 Voet, D. and Voet, J. G.)

*The original cholesterol structure proposed by Windaus and Wieland was incorrect. This became evident in 1932 as a result of X-ray diffraction studies done by the British physicist J. D. Bernal. By the end of 1932, however, English scientists, and Wieland himself, using Bernal's results, were able to outline the correct structure of cholesterol.

Certain compounds related to steroids and derived from plants are now known to lower total blood cholesterol when used in dietary forms approved by the FDA. Called phytostanols and phytosterols, these patented compounds act by inhibiting intestinal absorption of dietary cholesterol. They are marketed as food in the form of edible spreads. An example of a phytostanol is shown here.

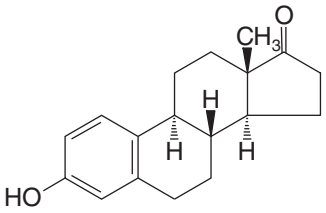


A phytostanol ester (β-sitostanol, R = fatty acid)

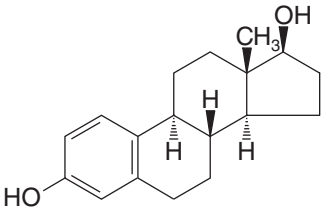
23.4C Sex Hormones

The sex hormones can be classified into three major groups: (1) the female sex hormones, or **estrogens;** (2) the male sex hormones, or **androgens;** and (3) the pregnancy hormones, or **progestins.**

The first sex hormone to be isolated was an estrogen, *estrone*. Working independently, Adolf Butenandt (in Germany at the University of Göttingen) and Edward Doisy (in the United States at St. Louis University) isolated estrone from the urine of pregnant women. They published their discoveries in 1929. Later, Doisy was able to isolate the much more potent estrogen, *estradiol*. In this research Doisy had to extract *4 tons* of sow ovaries in order to obtain just 12 mg of estradiol. Estradiol, it turns out, is the true female sex hormone, and estrone is a metabolized form of estradiol that is excreted.



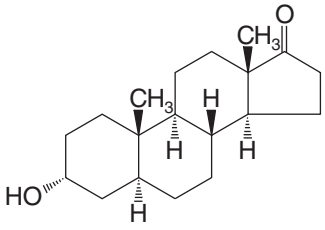
Estrone
[3-hydroxy-1,3,5(10)-estratrien-17-one]



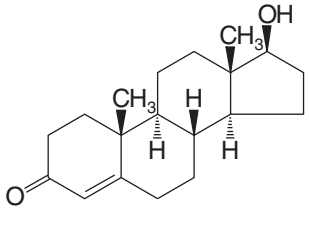
Estradiol
[1,3,5(10)-estratriene-3,17β-diol]

Estradiol is secreted by the ovaries and promotes the development of the secondary female characteristics that appear at the onset of puberty. Estrogens also stimulate the development of the mammary glands during pregnancy and induce estrus (heat) in animals.

In 1931, Butenandt and Kurt Tscherning isolated the first androgen, *androsterone*. They were able to obtain 15 mg of this hormone by extracting approximately 15,000 L of male urine. Soon afterward (in 1935), Ernest Laqueur (in Holland) isolated another male sex hormone, *testosterone*, from bull testes. It soon became clear that testosterone is the true male sex hormone and that androsterone is a metabolized form of testosterone that is excreted in the urine.



Androsterone
(3α-hydroxy-5 α-androstan-17-one)



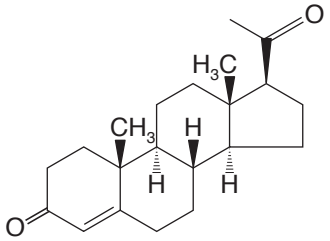
Testosterone
(17β-hydroxy-4-androsten-3-one)

Testosterone, secreted by the testes, is the hormone that promotes the development of secondary male characteristics: the growth of facial and body hair, the deepening of the voice, muscular development, and the maturation of the male sex organs.

Testosterone and estradiol, then, are the chemical compounds from which "maleness" and "femaleness" are derived. It is especially interesting to examine their structural formulas and see how very slightly these two compounds differ. Testosterone has an angular methyl group at the A,B ring junction that is missing in estradiol. Ring A of estradiol is a benzene ring and, as a result, estradiol is a phenol. Ring A of testosterone contains an α,β-unsaturated keto group.

Review Problem 23.9

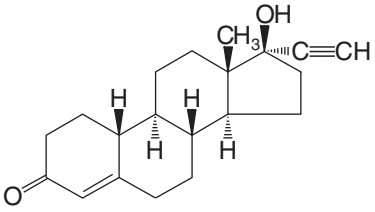
The estrogens (estrone and estradiol) are easily separated from the androgens (androsterone and testosterone) on the basis of one of their chemical properties. What is the property, and how could such a separation be accomplished?



Progesterone
(4-pregnene-3,20-dione)

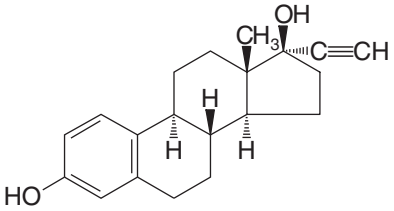
Progesterone is the most important *progestin* (pregnancy hormone). After ovulation occurs, the remnant of the ruptured ovarian follicle (called the *corpus luteum*) begins to secrete progesterone. This hormone prepares the lining of the uterus for implantation of the fertilized ovum, and continued progesterone secretion is necessary for the completion of pregnancy. (Progesterone is secreted by the placenta after secretion by the corpus luteum declines.)

Progesterone *also suppresses ovulation*, and it is the chemical agent that apparently accounts for the fact that pregnant women do not conceive again while pregnant. It was this observation that led to the search for synthetic progestins that could be used as oral contraceptives. (Progesterone itself requires very large doses to be effective in suppressing ovulation when taken orally because it is degraded in the intestinal tract.) A number of such compounds have been developed and are now widely used. In addition to norethynodrel (see Review Problem 23.7), another widely used synthetic progestin is its double-bond isomer, *norethindrone*:



Norethindrone
(17α-ethynyl-17β-hydroxy-4-estren-3-one)

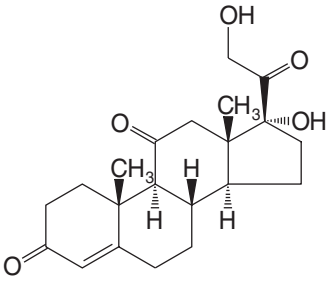
Synthetic estrogens have also been developed, and these are often used in oral contraceptives in combination with synthetic progestins. A very potent synthetic estrogen is the compound called *ethynylestradiol* or *novestrol*:



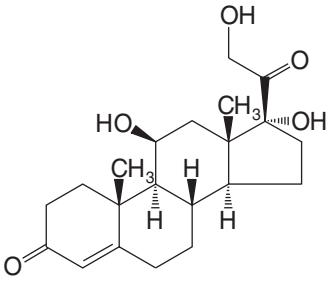
Ethynylestradiol
[17α-ethynyl-1,3,5(10)-estratriene-3,17β-diol]

23.4D Adrenocortical Hormones

At least 28 different hormones have been isolated from the adrenal cortex, part of the adrenal glands that sit on top of the kidneys. Included in this group are the following two steroids:



Cortisone
(17α,21-dihydroxy-4-pregnene-3,11,20-trione)



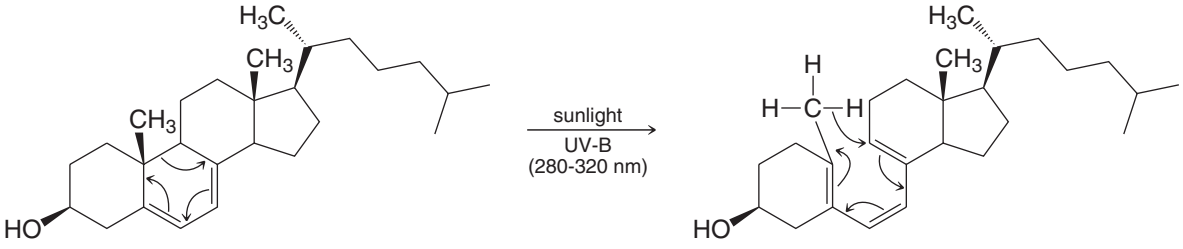
Cortisol
(11β,17α,21-trihydroxy-4-pregnene-3,20-dione)

Most of the adrenocortical steroids have an oxygen function at position 11 (a keto group in cortisone, for example, and a β-hydroxyl in cortisol). Cortisol is the major hormone synthesized by the human adrenal cortex.

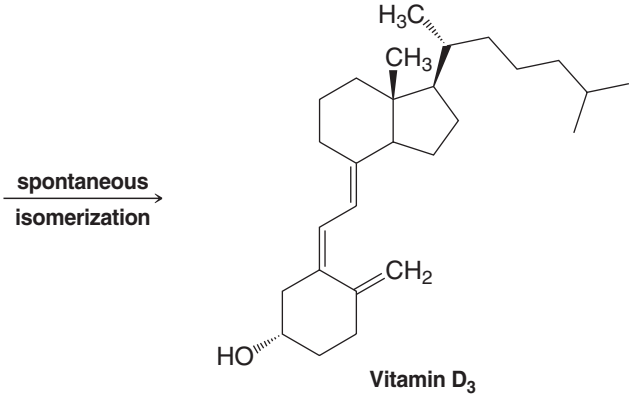
The adrenocortical steroids are apparently involved in the regulation of a large number of biological activities, including carbohydrate, protein, and lipid metabolism; water and electrolyte balance; and reactions to allergic and inflammatory phenomena. Recognition, in 1949, of the anti-inflammatory effect of cortisone and its usefulness in the treatment of rheumatoid arthritis led to extensive research in this area. Many 11-oxygenated steroids are now used in the treatment of a variety of disorders ranging from Addison's disease to asthma and skin inflammations.

23.4E D Vitamins

The demonstration, in 1919, that sunlight helped cure rickets—a childhood disease characterized by poor bone growth—began a search for a chemical explanation. Subsequent investigations showed that D vitamins were involved, and eventually it became known that one of several D vitamins, called vitamin D_3, is the curative factor. Vitamin D_3 is formed in the skin from 7-dehydrocholesterol by two reactions. In the first reaction (below),



7-Dehydrocholesterol **pre-vitamin D_3**



Vitamin D_3

ultraviolet light in the UV-B range (280–320 nm, which can penetrate the epidermal layer) brings about a 6-electron conrotatory electrocyclic reaction (see Special Topic H, *WileyPLUS*) to produce pre-vitamin D_3. Following this, a spontaneous isomerization (by way of a [1,7] sigmatropic hydride shift) produces vitamin D_3 itself.

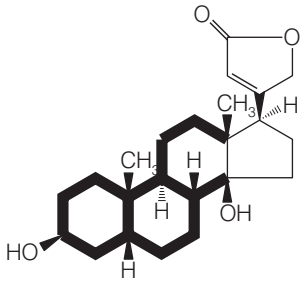
Vitamin D_3 is required for good health because it is essential in the process by which calcium (as Ca^{2+}) is absorbed from the intestine so as to allow for proper bone growth.

Various factors can cause a deficiency of sunlight and therefore of vitamin D_3, including one's geographic latitude and the season of the year. Sunlight levels are lower in extreme northern and southern latitudes, and are much lower in winter, so much so that for these conditions dietary guidelines in many countries call for supplemental D_3 for children and older persons. Other factors that can affect vitamin D_3 production in the skin are skin coloration, cloud cover, and the use of sunscreens.

23.4F Other Steroids

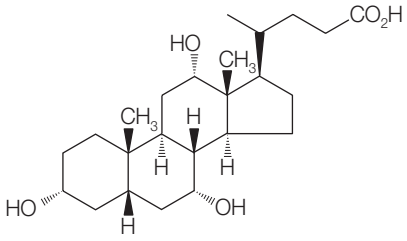
The structures, sources, and physiological properties of a number of other important steroids are given in Table 23.4.

TABLE 23.4 Other Important Steroids



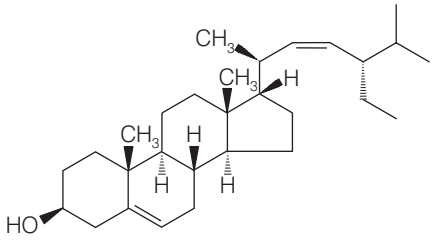
Digitoxigenin

Digitoxigenin is a cardiac aglycone that can be isolated by hydrolysis of digitalis, a pharmaceutical that has been used in treating heart disease since 1785. In digitalis, sugar molecules are joined in acetal linkages to the 3-OH group of the steroid. In small doses digitalis strengthens the heart muscle; in larger doses it is a powerful heart poison. The aglycone has only about one-fortieth the activity of digitalis.



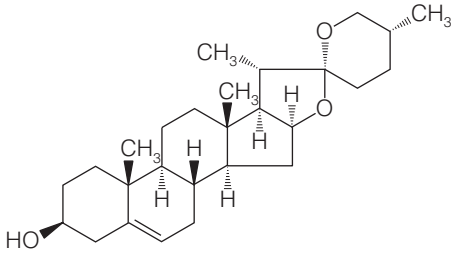
Cholic acid

Cholic acid is the most abundant acid obtained from the hydrolysis of human or ox bile. Bile is produced by the liver and stored in the gallbladder. When secreted into the small intestine, bile emulsifies lipids by acting as a soap. This action aids in the digestive process.



Stigmasterol

Stigmasterol is a widely occurring plant steroid that is obtained commercially from soybean oil. β-Sitostanol (a phytostanol, esters of which inhibit dietary cholesterol absorption) has the same formula except that it is saturated (C5 hydrogen is α).

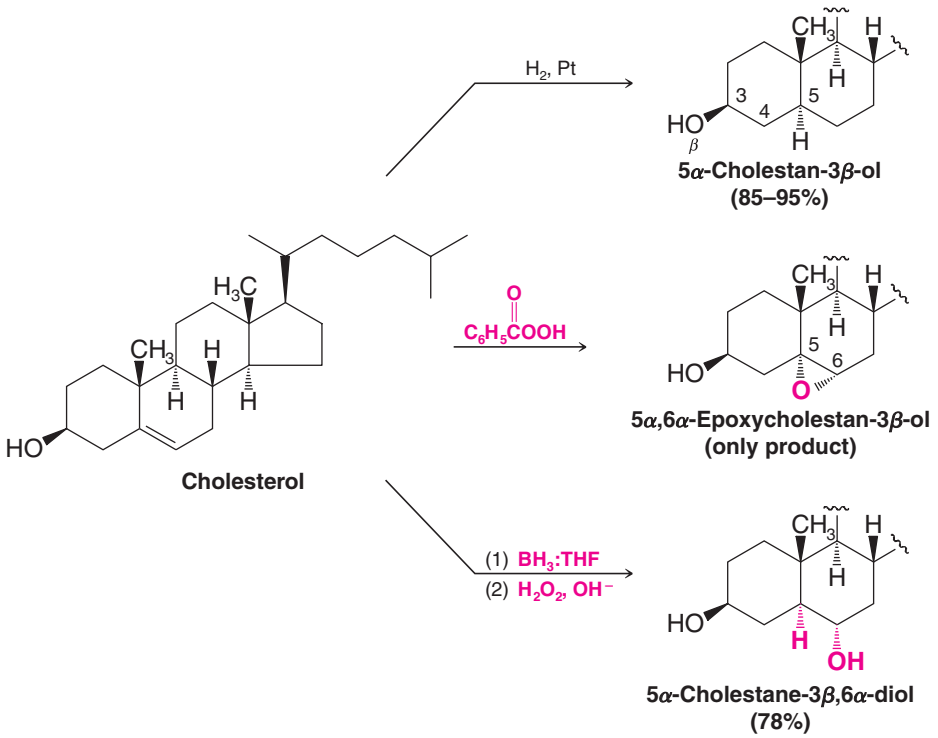


Diosgenin

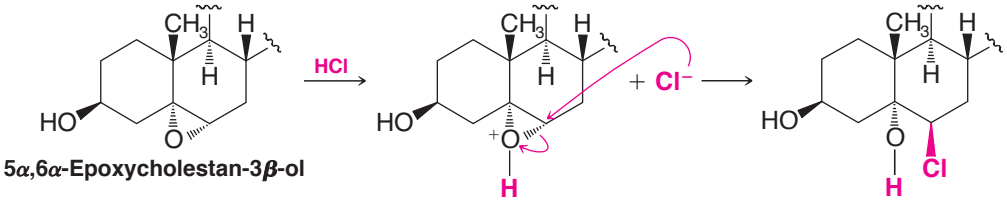
Diosgenin is obtained from a Mexican vine, *cabeza de negro*, genus *Dioscorea*. It is used as the starting material for a commercial synthesis of cortisone and sex hormones.

23.4G Reactions of Steroids

Steroids undergo all of the reactions that we might expect of molecules containing double bonds, hydroxyl groups, keto groups, and so on. While the stereochemistry of steroid reactions can be quite complex, it is often strongly influenced by the steric hindrance presented at the β face of the molecule by the angular methyl groups. Many reagents react preferentially at the relatively unhindered α face, especially when the reaction takes place at a functional group very near an angular methyl group and when the attacking reagent is bulky. Examples that illustrate this tendency are shown in the reactions below:



When the epoxide ring of 5α,6α-epoxycholestan-3β-ol (see the following reaction) is opened, attack by chloride ion must occur from the β face, but it takes place at the more open 6 position. Notice that the 5 and 6 substituents in the product are *diaxial* (Section 8.13):



Review Problem 23.10

Show how you might convert cholesterol into each of the following compounds:

(a) 5α,6β-Dibromocholestan-3β-ol

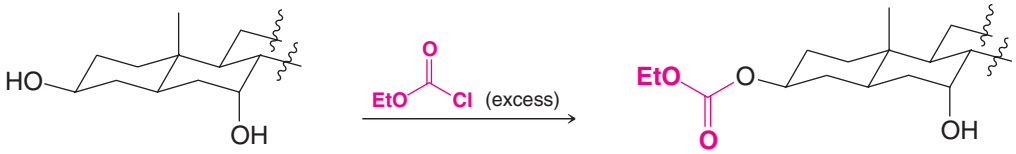
(b) Cholestane-3β,5α,6β-triol

(c) 5α-Cholestan-3-one

(d) 6α-Deuterio-5α-cholestan-3β-ol

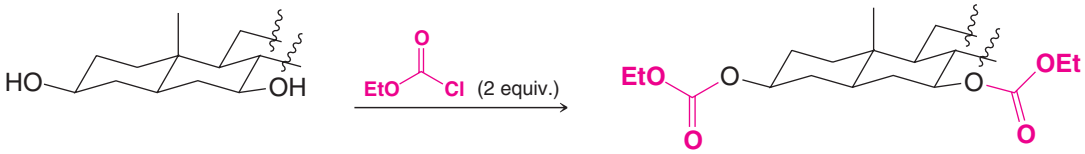
(e) 6β-Bromocholestane-3β,5α-diol

The relative openness of equatorial groups (when compared to axial groups) also influences the stereochemical course of steroid reactions. When 5α-cholestane-3β,7α-diol (see the following reaction) is treated with excess ethyl chloroformate (EtOCOCl), only the equatorial 3β-hydroxyl becomes esterified. The axial 7α-hydroxyl is unaffected by the reaction:



5α-Cholestane-3β,7α-diol (only product)

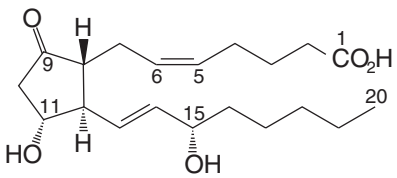
By contrast, treating 5α-cholestane-3β,7β-diol with excess ethyl chloroformate esterifies both hydroxyl groups. In this instance both groups are equatorial:



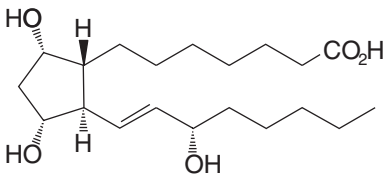
5α-Cholestane-3β,7β-diol (only product)

23.5 Prostaglandins

One very active area of research has concerned a group of lipids called **prostaglandins**. Prostaglandins are C_{20} carboxylic acids that contain a five-membered ring, at least one double bond, and several oxygen-containing functional groups. Two of the most biologically active prostaglandins are prostaglandin E_2 and prostaglandin $F_{1\alpha}$:



Prostaglandin E_2 (PGE_2)



Prostaglandin $F_{1\alpha}$ ($PGF_{1\alpha}$)

Helpful Hint

These names for the prostaglandins are abbreviated designations used by workers in the field; systematic names are seldom used for prostaglandins.

Prostaglandins of the E type have a carbonyl group at C9 and a hydroxyl group at C11; those of the F type have hydroxyl groups at both positions. Prostaglandins of the 2 series have a double bond between C5 and C6; in the 1 series this bond is a single bond.

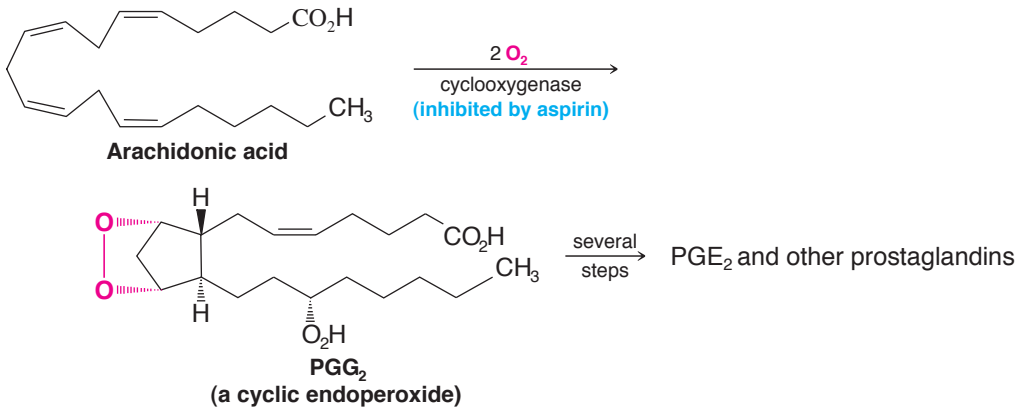
First isolated from seminal fluid, prostaglandins have since been found in almost all animal tissues. The amounts vary from tissue to tissue but are almost always very small. Most prostaglandins have powerful physiological activity, however, and this activity covers a broad spectrum of effects. Prostaglandins are known to affect heart rate, blood pressure, blood clotting, conception, fertility, and allergic responses.

The finding that prostaglandins can prevent formation of blood clots has great clinical significance, because heart attacks and strokes often result from the formation of abnormal clots in blood vessels. An understanding of how prostaglandins affect the formation of clots may lead to the development of drugs to prevent heart attacks and strokes.

The biosynthesis of prostaglandins of the 2 series begins with a C_{20} polyenoic acid, arachidonic acid, an omega-6 fatty acid. (Synthesis of prostaglandins of the 1 series begins

The 1982 Nobel Prize in Physiology or Medicine was awarded to S. K. Bergström and B. I. Samuelsson (Karolinska Institute, Stockholm, Sweden) and to J. R. Vane (Wellcome Foundation, Beckenham, England) for their work on prostaglandins.

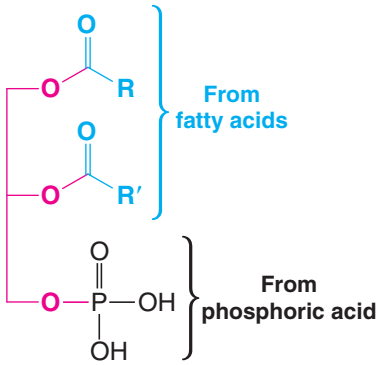
with a fatty acid with one fewer double bond.) The first step requires two molecules of oxygen and is catalyzed by an enzyme called *cyclooxygenase*:



The involvement of prostaglandins in allergic and inflammatory responses has also been of special interest. Some prostaglandins induce inflammation; others relieve it. The most widely used anti-inflammatory drug is ordinary aspirin (see Section 21.8). Aspirin blocks the synthesis of prostaglandins from arachidonic acid, apparently by acetylating the enzyme cyclooxygenase, thus rendering it inactive (see the previous reaction). This reaction may represent the origin of aspirin's anti-inflammatory properties. Another prostaglandin (PGE_1) is a potent fever-inducing agent (pyrogen), and aspirin's ability to reduce fever may also arise from its inhibition of prostaglandin synthesis.

23.6 Phospholipids and Cell Membranes

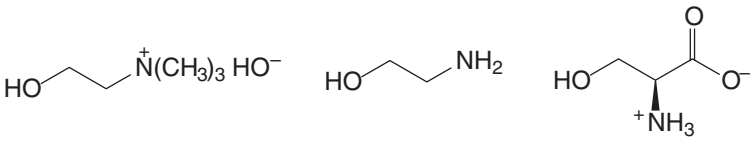
Another large class of lipids are those called **phospholipids**. Most phospholipids are structurally derived from a glycerol derivative known as a *phosphatidic acid*. In a phosphatidic acid, two hydroxyl groups of glycerol are joined in ester linkages to fatty acids and one terminal hydroxyl group is joined in an ester linkage to *phosphoric acid*:



A phosphatidic acid
(a diacylglyceryl phosphate)

23.6A Phosphatides

In *phosphatides*, the phosphate group of a phosphatidic acid is bound through another phosphate ester linkage to one of the following nitrogen-containing compounds:

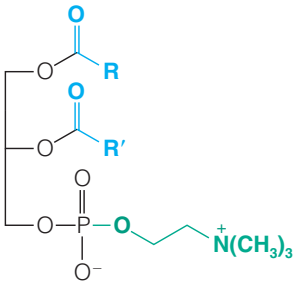


Choline | **2-Aminoethanol (ethanolamine)** | **L-Serine**

The most important phosphatides are the **lecithins, cephalins, phosphatidylserines,** and **plasmalogens** (a phosphatidyl derivative). Their general structures are shown in Table 23.5.

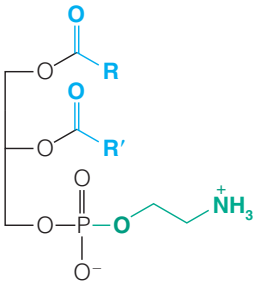
TABLE 23.5 Phosphatides

Lecithins



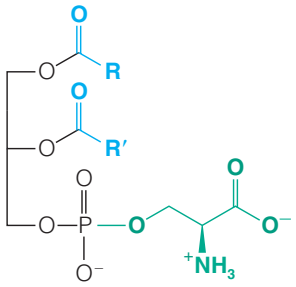
(from choline)
R is saturated and R′ is unsaturated.

Cephalins



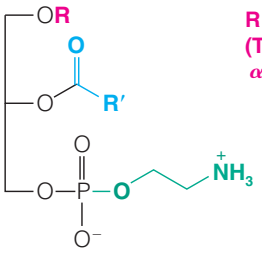
(from 2-aminoethanol)

Phosphatidylserines



(from L-serine)
R is saturated and R′ is unsaturated.

Plasmalogens



R is —CH═CH(CH$_2$)$_n$CH$_3$ (This linkage is that of an α,β-unsaturated ether.)

(from 2-aminoethanol) or —OCH$_2$CH$_2$N$^+$(CH$_3$)$_3$ (from choline)
R′ is an unsaturated fatty acid.

Phosphatides resemble soaps and detergents in that they are molecules having both polar and nonpolar groups (Fig. 23.8*a*). Like soaps and detergents, too, phosphatides "dissolve" in aqueous media by forming micelles. There is evidence that in biological systems the preferred micelles consist of three-dimensional arrays of "stacked" bimolecular micelles (Fig. 23.8*b*) that are better described as **lipid bilayers**.

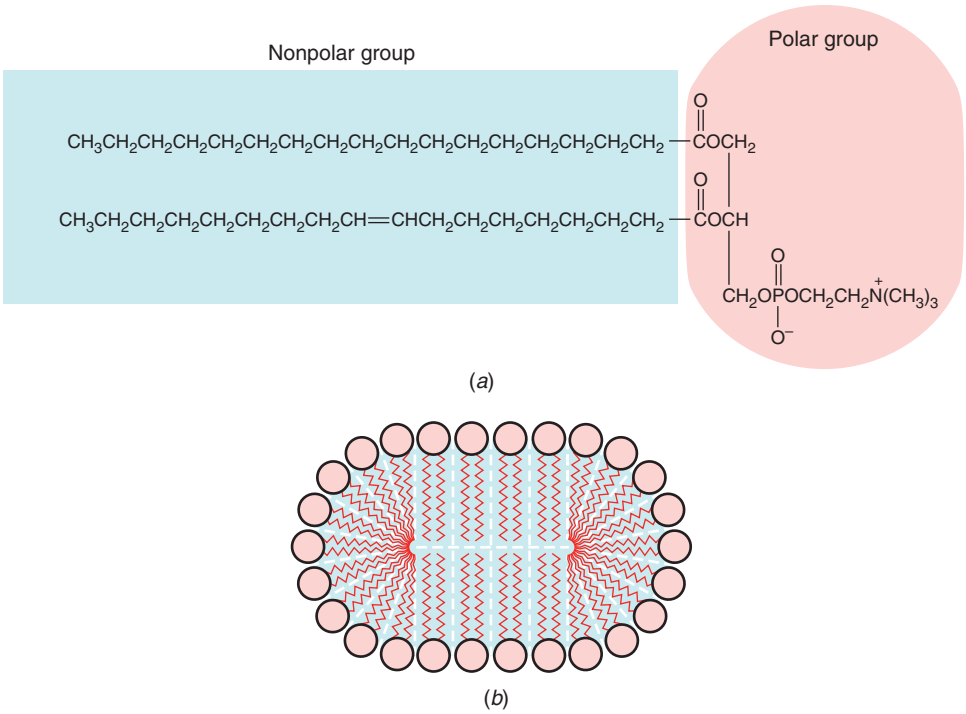


Figure 23.8 (*a*) Polar and nonpolar sections of a phosphatide. (*b*) A phosphatide micelle or lipid bilayer.

The hydrophilic and hydrophobic portions of phosphatides make them perfectly suited for one of their most important biological functions: They form a portion of a structural unit that creates an interface between an organic and an aqueous environment. This structure (Fig. 23.9) is located in cell walls and membranes where phosphatides are often found associated with proteins and glycolipids (Section 23.6B).

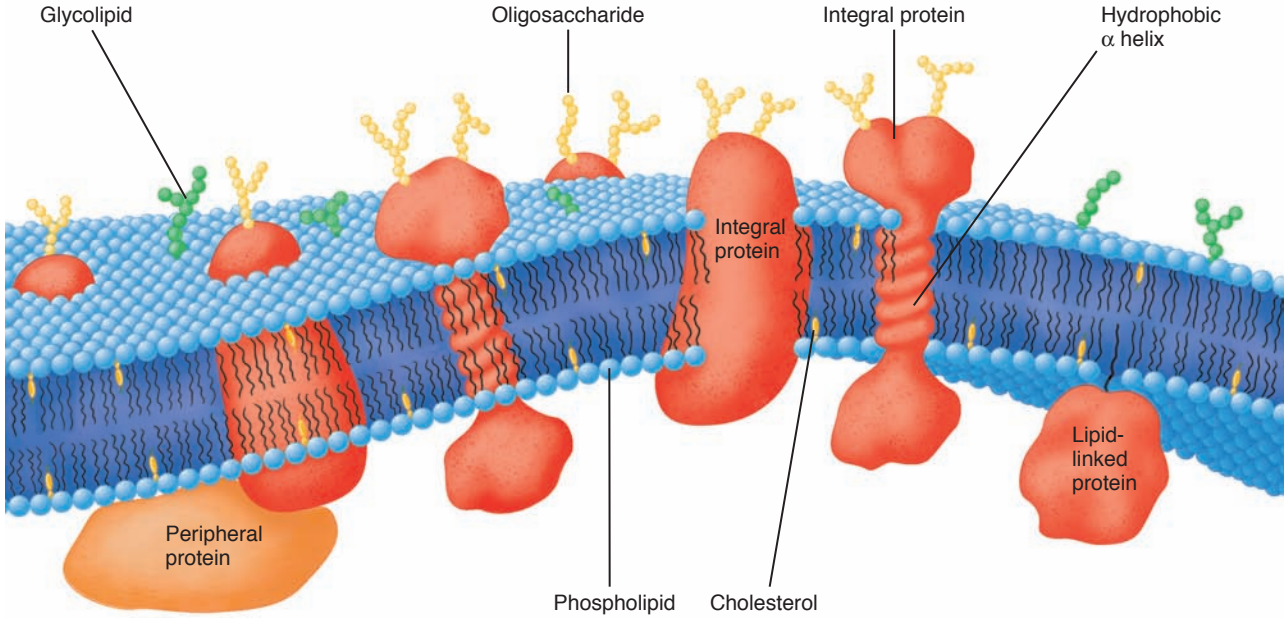
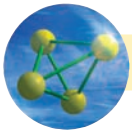


Figure 23.9 A schematic diagram of a plasma membrane. Integral proteins (*red-orange*), shown for clarity in much greater proportion than they are found in actual biological membranes, and cholesterol (*yellow*) are embedded in a bilayer composed of phospholipids (*blue spheres with two wiggly tails*). The carbohydrate components of glycoproteins (*yellow beaded chains*) and glycolipids (*green beaded chains*) occur only on the external face of the membrane. (Reprinted with permission of John Wiley & Sons, Inc., from Voet, D.; Voet, J. G.; Pratt, C., *Fundamentals of Biochemistry, Life at the Molecular Level*; © 1999 Voet, D. and Voet, J. G.)

Under suitable conditions all of the ester (and ether) linkages of a phosphatide can be hydrolyzed. What organic compounds would you expect to obtain from the complete hydrolysis of (see Table 23.5) **(a)** a lecithin, **(b)** a cephalin, and **(c)** a choline-based plasmalogen? [*Note:* Pay particular attention to the fate of the α,β-unsaturated ether in part (c).

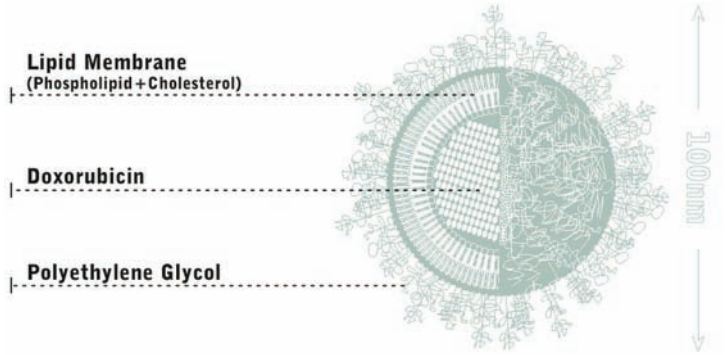
Review Problem 23.11



THE CHEMISTRY OF . . .

STEALTH® Liposomes for Drug Delivery

The anticancer drug Doxil (doxorubicin) has been packaged in STEALTH® liposomes that give each dose of the drug extended action in the body. During manufacture of the drug it is ensconced in microscopic bubbles (vesicles) formed by a phospholipid bilayer and then given a special coating that masks it from the immune system. Ordinarily, a foreign particle such as this would be attacked by cells of the immune system and degraded, but a veil of polyethylene glycol oligomers on the liposome surface masks it from detection. Because of this coating, the STEALTH® liposome circulates through the body and releases its therapeutic contents over a period of time significantly greater than the lifetime for circulation of the undisguised drug. Coatings like those used for STEALTH® liposomes may also be able to reduce the toxic side effects of some drugs. Furthermore, by attaching specific cell recognition "marker" molecules to the polymer, it may be possible to focus binding of the liposomes specifically to cells of a targeted tissue. One might be tempted to call a targeted liposome a "smart stealth liposome."



A STEALTH® liposome carrying the drug doxorubicin. (© and courtesy of ALZA Corporation.)

23.6B Derivatives of Sphingosine

Another important group of lipids is derived from **sphingosine;** the derivatives are called **sphingolipids.** Two sphingolipids, a typical *sphingomyelin* and a typical *cerebroside*, are shown in Fig. 23.10.

On hydrolysis, sphingomyelins yield sphingosine, choline, phosphoric acid, and a C_{24} fatty acid called lignoceric acid. In a sphingomyelin this last component is bound to the $-NH_2$ group of sphingosine. The sphingolipids do not yield glycerol when they are hydrolyzed.

The cerebroside shown in Fig. 23.10 is an example of a **glycolipid.** Glycolipids have a polar group that is contributed by a *carbohydrate*. They do not yield phosphoric acid or choline when they are hydrolyzed.

The sphingolipids, together with proteins and polysaccharides, make up **myelin,** the protective coating that encloses nerve fibers or **axons.** The axons of nerve cells carry electrical nerve impulses. Myelin has a function relative to the axon similar to that of the insulation on an ordinary electric wire (see the chapter opening vignette).

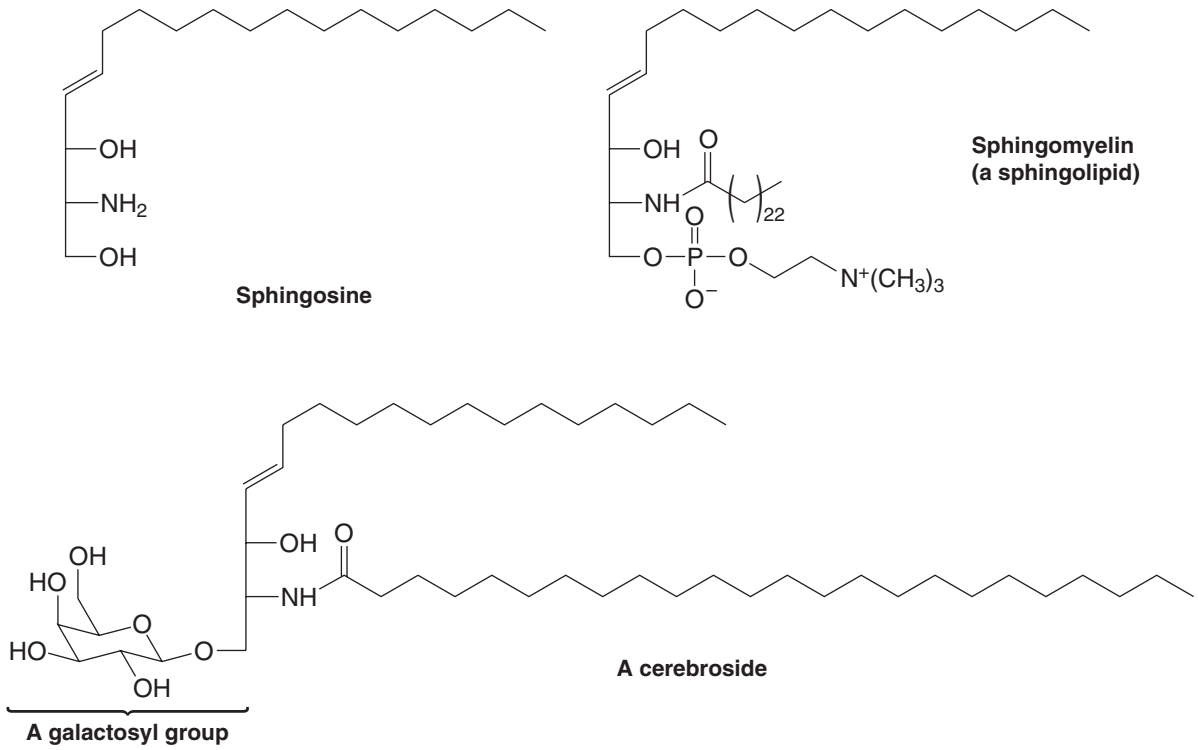
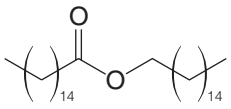


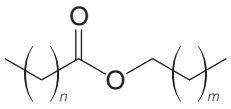
Figure 23.10 A sphingosine and two sphingolipids.

23.7 *Waxes*

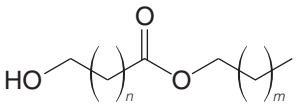
Most **waxes** are esters of long-chain fatty acids and long-chain alcohols. Waxes are found as protective coatings on the skin, fur, and feathers of animals and on the leaves and fruits of plants. Several esters isolated from waxes are the following:



Cetyl palmitate
(from spermaceti)



***n* = 24 or 26; *m* = 28 or 30**
(from beeswax)



***n* = 16–28; *m* = 30 or 32**
(from carnauba wax)

Summary of Reactions of Lipids

The reactions of lipids represent many reactions that we have studied in previous chapters, especially reactions of carboxylic acids, alkenes, and alcohols. Ester hydrolysis (e.g., saponification) liberates fatty acids and glycerol from triacylglycerols. The carboxylic acid group of a fatty acid can be reduced, converted to an activated acyl derivative such as an acyl chloride, or converted to an ester or amide. Alkene functional groups in unsaturated fatty acids can be hydrogenated, hydrated, halogenated, hydrohalogenated, converted to a vicinal diol or epoxide, or cleaved by oxidation reactions. Alcohol functional groups in lipids such as terpenes, steroids, and prostaglandins can be alkylated, acylated, oxidized, or used in elimination reactions. All of these are reactions we have studied previously in the context of smaller molecules.

Key Terms and Concepts

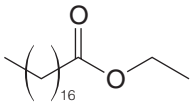
The key terms and concepts that are highlighted in **bold, blue text** within the chapter are defined in the glossary (at the back of the book) and have hyperlinked definitions in the accompanying ***WileyPLUS*** course (www.wileyplus.com).

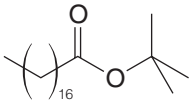
Problems

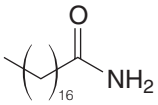
Note to Instructors: Many of the homework problems are available for assignment via ***WileyPLUS***, an online teaching and learning solution.

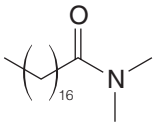
GENERAL REACTIONS

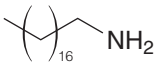
23.12 How would you convert stearic acid, $CH_3(CH_2)_{16}CO_2H$, into each of the following?

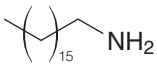
(a) Ethyl stearate,  (two ways)

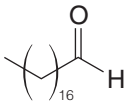
(b) *tert*-Butyl stearate, 

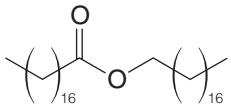
(c) Stearamide, 

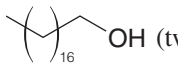
(d) *N,N*-Dimethylstearamide, 

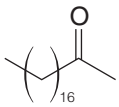
(e) Octadecylamine, 

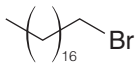
(f) Heptadecylamine, 

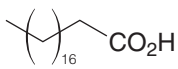
(g) Octadecanal, 

(h) Octadecyl stearate, 

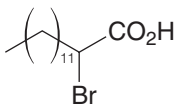
(i) 1-Octadecanol,  (two ways)

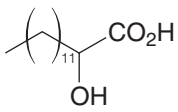
(j) 2-Nonadecanone, 

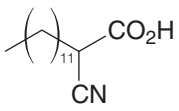
(k) 1-Bromooctadecane, 

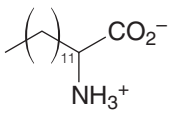
(l) Nonadecanoic acid, 

23.13 How would you transform tetradecanal into each of the following?

(a) 

(b) 

(c) 

(d) 

23.14 Using palmitoleic acid as an example and neglecting stereochemistry, illustrate each of the following reactions of the double bond:

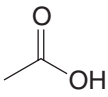
(a) Addition of bromine **(b)** Addition of hydrogen **(c)** Hydroxylation **(d)** Addition of HCl

23.15 When oleic acid is heated to 180–200°C (in the presence of a small amount of selenium), an equilibrium is established between oleic acid (33%) and an isomeric compound called elaidic acid (67%). Suggest a possible structure for elaidic acid.

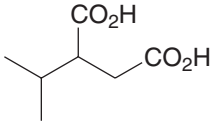
23.16 When limonene (Section 23.3) is heated strongly, it yields 2 mol of isoprene. What kind of reaction is involved here?

23.17 Gadoleic acid ($C_{20}H_{38}O_2$), a fatty acid that can be isolated from cod-liver oil, can be cleaved by hydroxylation and subsequent treatment with periodic acid to $CH_3(CH_2)_9CHO$ and $OHC(CH_2)_7CO_2H$. **(a)** What two stereoisomeric structures are possible for gadoleic acid? **(b)** What spectroscopic technique would make possible a decision as to the actual structure of gadoleic acid? **(c)** What peaks would you look for?

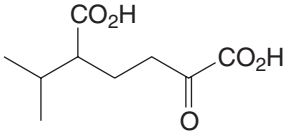
23.18 α-Phellandrene and β-phellandrene are isomeric compounds that are minor constituents of spearmint oil; they have the molecular formula $C_{10}H_{16}$. Each compound has a UV absorption maximum in the 230–270-nm range. On catalytic hydrogenation, each compound yields 1-isopropyl-4-methylcyclohexane. On vigorous oxidation with potassium permanganate, α-phellandrene yields



and



A similar oxidation of β-phellandrene yields



as the only isolable product. Propose structures for α- and β-phellandrene.

ROADMAP SYNTHESES

23.19 Vaccenic acid, a constitutional isomer of oleic acid, has been synthesized through the following reaction sequence:

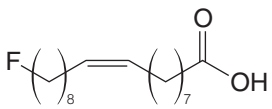
$$\text{1-Octyne} + NaNH_2 \xrightarrow[NH_3]{\text{liq.}} \mathbf{A}\ (C_8H_{13}Na) \xrightarrow{ICH_2(CH_2)_7CH_2Cl} \mathbf{B}\ (C_{17}H_{31}Cl) \xrightarrow{NaCN} \mathbf{C}\ (C_{18}H_{31}N) \xrightarrow{KOH,\ H_2O} \mathbf{D}\ (C_{18}H_{31}O_2K) \xrightarrow{H_3O^+}$$

$$\mathbf{E}\ (C_{18}H_{32}O_2) \xrightarrow[BaSO_4]{H_2,\ Pd} \text{vaccenic acid}\ (C_{18}H_{34}O_2)$$

Propose a structure for vaccenic acid and for the intermediates **A–E.**

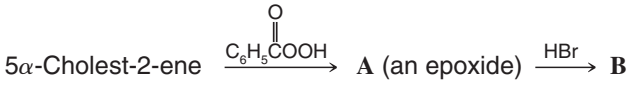
23.20 ω-Fluorooleic acid can be isolated from a shrub, *Dechapetalum toxicarium*, that grows in Africa. The compound is highly toxic to warm-blooded animals; it has found use as an arrow poison in tribal warfare, in poisoning enemy water supplies, and by witch doctors "for terrorizing the native population." Powdered fruit of the plant has been used as a rat poison; hence ω-fluorooleic acid has the common name "ratsbane." A synthesis of ω-fluorooleic acid is outlined here. Give structures for compounds **F–I:**

$$\text{1-Bromo-8-fluorooctane} + \text{sodium acetylide} \longrightarrow \mathbf{F}\ (C_{10}H_{17}F) \xrightarrow[(2)\ I(CH_2)_7Cl]{(1)\ NaNH_2} \mathbf{G}\ (C_{17}H_{30}FCl) \xrightarrow{NaCN} \mathbf{H}\ (C_{18}H_{30}NF) \xrightarrow[(2)\ H_3O^+]{(1)\ KOH} \mathbf{I}\ (C_{18}H_{31}O_2F) \xrightarrow[Ni_2B\ (P\text{-}2)]{H_2}$$



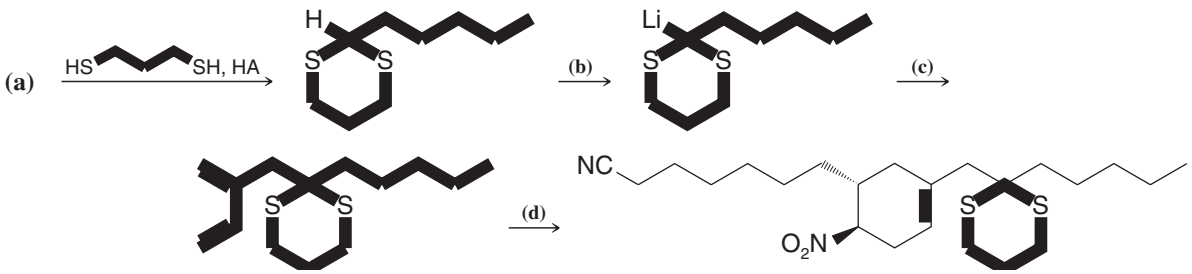
ω-Fluorooleic acid
(46% yield, overall)

23.21 Give formulas and names for compounds **A** and **B:**

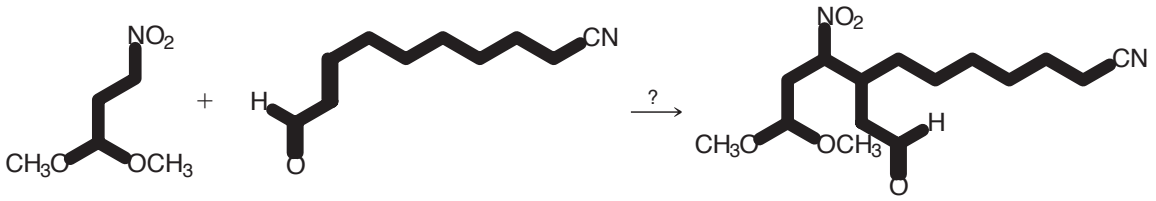


(*Hint*: **B** is not the most stable stereoisomer.)

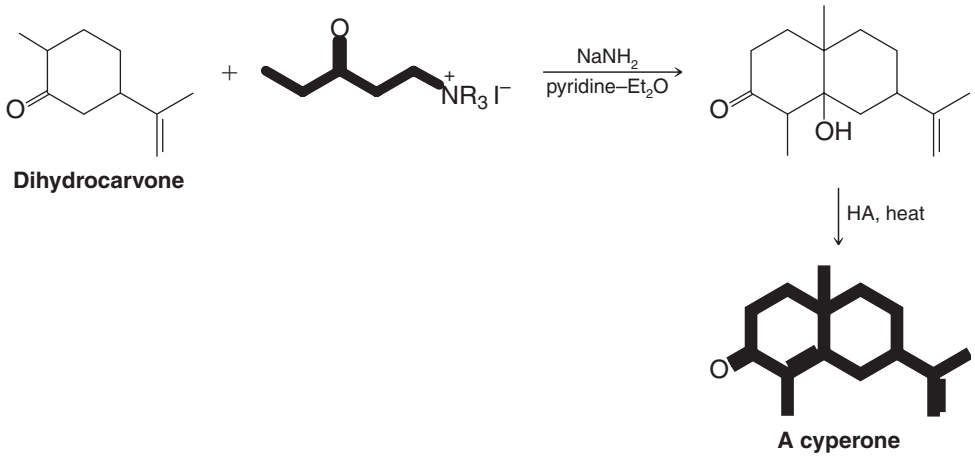
23.22 The initial steps of a laboratory synthesis of several prostaglandins reported by E. J. Corey (Section 7.16B) and co-workers in 1968 are outlined here. Supply each of the missing reagents:



(e) The initial step in another prostaglandin synthesis is shown in the following reaction. What kind of reaction—and catalyst—is needed here?



23.23 A useful synthesis of sesquiterpene ketones, called *cyperones*, was accomplished through a modification of the following Robinson annulation procedure (Section 19.7B).



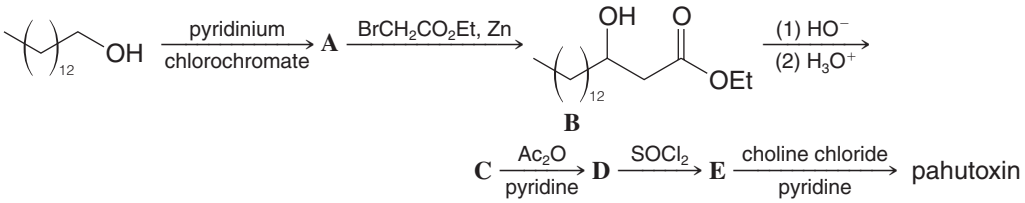
Dihydrocarvone

A cyperone

Write a mechanism that accounts for each step of this synthesis.

Challenge Problems

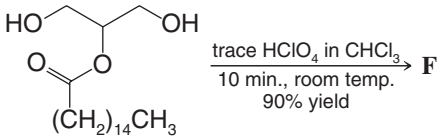
23.24 A Hawaiian fish called the pahu or boxfish (*Ostracian lentiginosus*) secretes a toxin that kills other fish in its vicinity. The active agent in the secretion was named pahutoxin by P. J. Scheuer, and it was found by D. B. Boylan and Scheuer to contain an unusual combination of lipid moieties. To prove its structure, they synthesized it by this route:



Compound	Selected Infrared Absorption Bands (cm^{-1})
A	1725
B	3300 (broad), 1735
C	3300–2500 (broad), 1710
D	3000–2500 (broad), 1735, 1710
E	1800, 1735
Pahutoxin	1735

What are the structures of **A, C, D,** and **E** and of pahutoxin?

23.25 The reaction illustrated by the equation below is a very general one that can be catalyzed by acid, base, and some enzymes. It therefore needs to be taken into consideration when planning syntheses that involve esters of polyhydroxy substances like glycerol and sugars:



Spectral data for **F:**

MS (*m/z*): (after trimethylsilylation): 546, 531

IR (cm^{-1}, in CCl_4 solution): 3200 (broad), 1710

^{1}H NMR (δ) (after exchange with D_2O): 4.2 (d), 3.9 (m), 3.7 (d), 2.2 (t), and others in the range 1.7 to 1

^{13}C NMR (δ): 172 (C), 74 (CH), 70 (CH_2), 67 (CH_2), 39 (CH_2), and others in the range 32 to 14

(a) What is the structure of product **F**?

(b) The reaction is intramolecular. Write a mechanism by which it probably occurs.

Learning Group Problems

1. Olestra is a fat substitute patented by Procter and Gamble that mimics the taste and texture of triacylglycerols (see "The Chemistry of . . . Olestra and Other Fat Substitutes" in Section 23.2B). It is calorie-free because it is neither hydrolyzed by digestive enzymes nor absorbed by the intestines but instead is passed directly through the body unchanged. The FDA has approved olestra for use in a variety of foods, including potato chips and other snack foods that typically have a high fat content. It can be used in both the dough and the frying process.

(a) Olestra consists of a mixture of sucrose fatty acid esters (unlike triacylglycerols, which are glycerol esters of fatty acids). Each sucrose molecule in olestra is esterified with six to eight fatty acids. (One undesirable aspect of olestra is that it sequesters fat-soluble vitamins needed by the body, due to its high lipophilic character.) Draw the structure of a specific olestra molecule comprising six different naturally occurring fatty acids esterified to any of the available positions on sucrose. Use three saturated fatty acids and three unsaturated fatty acids.

(b) Write reaction conditions that could be used to saponify the esters of the olestra molecule you drew and give IUPAC and common names for each of the fatty acids that would be liberated on saponification.

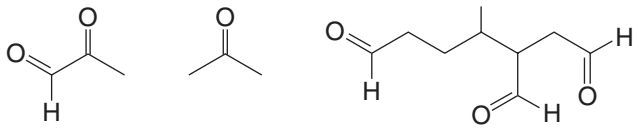
(c) Olestra is made by sequential transesterification processes. The first transesterification involves reaction of methanol under basic conditions with natural triacylglycerols from cottonseed or soybean oil (chain lengths of C_8–C_{22}). The second transesterification involves reaction of these fatty acid methyl esters with sucrose to form olestra. Write one example reaction, including its mechanism, for each of these transesterification processes used in the synthesis of olestra. Start with any triacylglycerol having fatty acids like those incorporated into olestra.

2. The biosynthesis of fatty acids is accomplished two carbons at a time by an enzyme complex called fatty acid synthetase. The biochemical reactions involved in fatty acid synthesis are described in Special Topic E (*WileyPLUS*). Each of these biochemical reactions has a counterpart in synthetic reactions you have studied. Consider the biochemical reactions involved in adding each $—CH_2CH_2—$ segment during fatty acid biosynthesis (those in Special Topic E that begin with acetyl-S-ACP and malonyl-S-ACP, and end with butyryl-S-ACP). Write laboratory synthetic reactions using reagents and conditions you have studied (not biosynthetic reactions) that would accomplish the same sequence of transformations (i.e., the condensation–decarboxylation, ketone reduction, dehydration, and alkene reduction steps).

3. A certain natural terpene produced peaks in its mass spectrum at *m/z* 204, 111, and 93 (among others). On the basis of this and the following information, elucidate the structure of this terpene. Justify each of your conclusions.

(a) Reaction of the unknown terpene with hydrogen in the presence of platinum under pressure results in a compound with molecular formula $C_{15}H_{30}$.

(b) Reaction of the terpene with ozone followed by dimethyl sulfide produces the following mixture of compounds (1 mol of each for each mole of the unknown terpene):



(c) After writing the structure of the unknown terpene, circle each of the isoprene units in this compound. To what class of terpenes does this compound belong (based on the number of carbons it contains)?

4. Draw the structure of a phospholipid (from any of the subclasses of phospholipids) that contains one saturated and one unsaturated fatty acid.

(a) Draw the structure of all of the products that would be formed from your phospholipid if it were subjected to complete hydrolysis (choose either acidic or basic conditions).

(b) Draw the structure of the product(s) that would be formed from reaction of the unsaturated fatty acid moiety of your phospholipid (assuming it had been released by hydrolysis from the phospholipid first) under each of the following conditions:

(i) Br_2 in CCl_4

(ii) OsO_4, followed by $NaHSO_3$

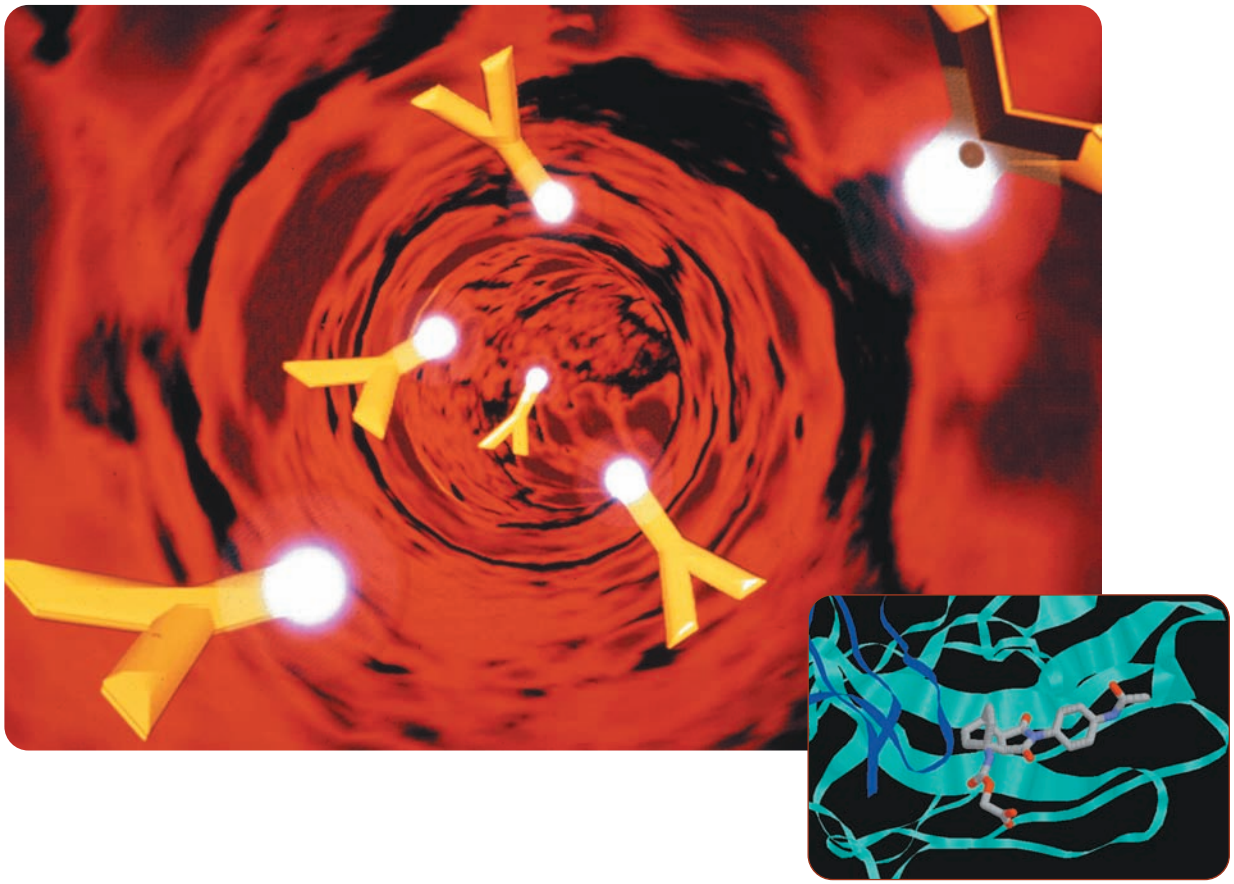
(iii) HBr

(iv) Hot alkaline $KMnO_4$, followed by H_3O^+

(v) $SOCl_2$, followed by excess CH_3NH_2

24

Amino Acids and Proteins



A synthetic Diels-Alderase catalytic antibody with a bound hapten.

Chemists are capitalizing on the natural adaptability of the immune system to create what we can fittingly call *designer catalysts*. These catalysts are *antibodies*—protein species usually produced by the immune system to capture and remove foreign agents but which, in this case, are elicited in a way that makes them able to catalyze chemical reactions.

The creation of the first catalytic antibodies by Richard A. Lerner and Peter G. Schultz (both of Scripps Research Institute) represented an ingenious union of principles relating to enzyme chemistry and the innate capabilities of the immune system. In some respects catalytic antibodies are like enzymes, the protein catalysts we have mentioned many times already and shall study further in this chapter. Unlike enzymes, however, catalytic antibodies can virtually be "made to order" for specific reactions by a marriage of chemistry and immunology. Examples include catalytic antibodies for Claisen rearrangements, Diels–Alder reactions (such as that shown in the molecular graphic above), ester hydrolyses, and aldol reactions. We shall consider how catalytic antibodies are produced in "The Chemistry of . . . Some Catalytic Antibodies" later in this chapter. Designer catalysts are indeed at hand.

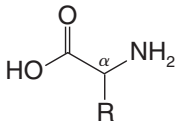
24.1 Introduction

The three groups of biological polymers are polysaccharides, proteins, and nucleic acids. We studied polysaccharides in Chapter 22 and saw that they function primarily as energy reserves, as biochemical labels on cell surfaces, and, in plants, as structural materials. When we study nucleic acids in Chapter 25, we shall find that they serve two major purposes: storage and transmission of information. Of the three groups of biopolymers, proteins have the most diverse functions. As enzymes and hormones, proteins catalyze and regulate the reactions that occur in the body; as muscles and tendons they provide the body with the means for movement; as skin and hair they give it an outer covering; as hemoglobin molecules they transfer all-important oxygen to its most remote corners; as antibodies they provide it with a means of protection against disease; and in combination with other substances in bone they provide it with structural support.

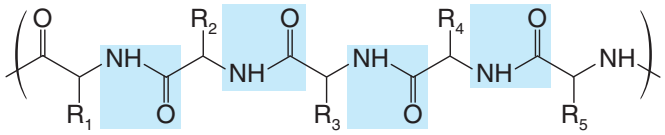
Given such diversity of functions, we should not be surprised to find that proteins come in all sizes and shapes. By the standard of most of the molecules we have studied, even small proteins have very high molecular weights. Lysozyme, an enzyme, is a relatively small protein and yet its molecular weight is 14,600. The molecular weights of most proteins are much larger. Their shapes cover a range from the globular proteins such as lysozyme and hemoglobin to the helical coils of α-keratin (hair, nails, and wool) and the pleated sheets of silk fibroin.

And yet, in spite of such diversity of size, shape, and function, all proteins have common features that allow us to deduce their structures and understand their properties. Later in this chapter we shall see how this is done.

- Proteins are **polyamides,** and their monomeric units are composed of about 20 different α-amino acids:



An α-amino acid
R is a side chain at the α carbon that determines the identity of the amino acid (Table 24.1).



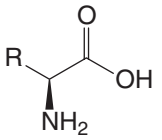
A portion of a protein molecule
Amide (peptide) linkages are shaded.
R_1 to R_5 may be any of the possible side chains.

- The exact sequence of the different α-amino acids along the protein chain is called the **primary structure** of the protein.

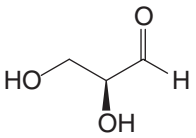
A protein's primary structure, as its name suggests, is of fundamental importance. For the protein to carry out its particular function, the primary structure must be correct. We shall see later that when the primary structure is correct, the protein's polyamide chain folds in particular ways to give it the shape it needs for its particular task.

- Folding of the polyamide chain gives rise to higher levels of complexity called the **secondary** and **tertiary structures** of the protein.
- **Quaternary structure** results when a protein contains an aggregate of more than one polyamide chain.
- Hydrolysis of proteins with acid or base yields a mixture of amino acids.

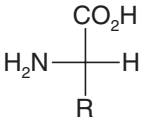
Although hydrolysis of naturally occurring proteins may yield as many as 22 different amino acids, the amino acids have an important structural feature in common: With the exception of glycine (whose molecules are achiral), almost all naturally occurring amino acids have the L configuration at the α carbon.* That is, they have the same relative configuration as L-glyceraldehyde:

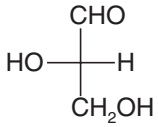


An L-α-amino acid [usually an (*S*)-α-amino acid]



L-Glyceraldehyde [(*S*)-glyceraldehyde]





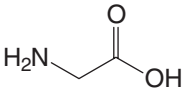
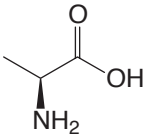
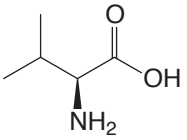
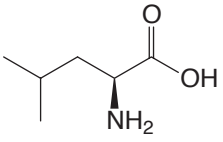
Fischer projections for an L-α-amino acid and L-glyceraldehyde

24.2 Amino Acids

24.2A Structures and Names

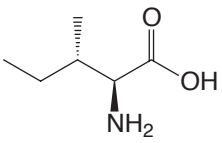
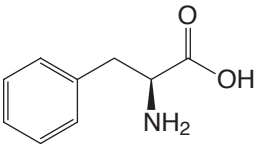
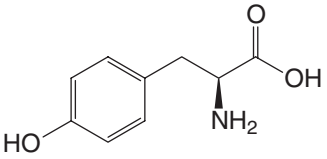
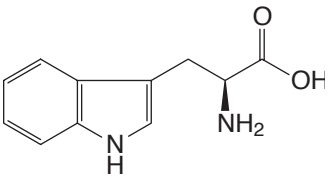
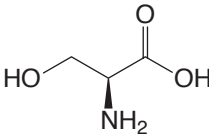
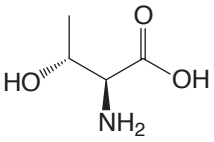
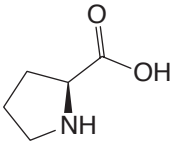
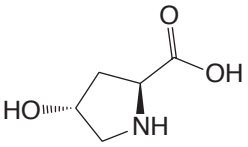
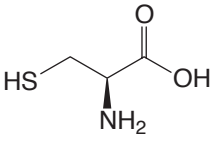
- The 22 α-amino acids that can be obtained from proteins can be subdivided into three different groups on the basis of the structures of their side chains, R. These are given in Table 24.1.

TABLE 24.1 L-Amino Acids Found in Proteins

Structure	Name	Abbreviations[a]	pK_{a_1} α-CO_2H	pK_{a_2} α-NH_3^+	pK_{a_3} R group	p*I*
Neutral Amino Acids						
	Glycine	G or Gly	2.3	9.6		6.0
	Alanine	A or Ala	2.3	9.7		6.0
	Valine[b]	V or Val	2.3	9.6		6.0
	Leucine[b]	L or Leu	2.4	9.6		6.0

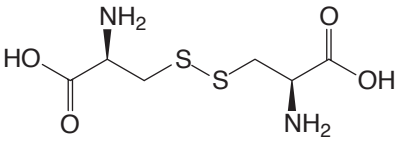
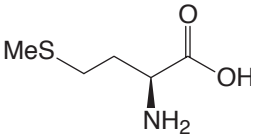
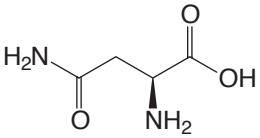
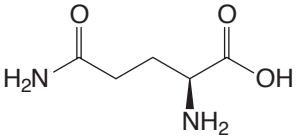
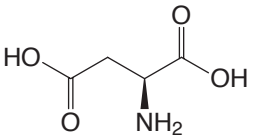
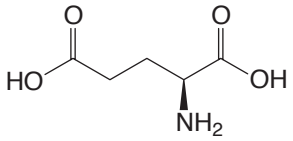
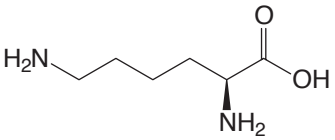
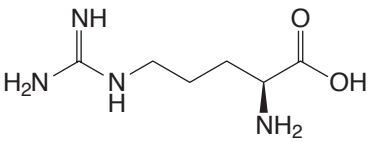
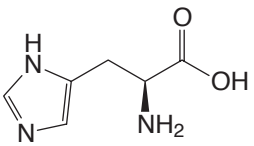
*Some D-amino acids have been obtained from the material comprising the cell walls of bacteria and by hydrolysis of certain antibiotics.

TABLE 24.1 CONTINUED

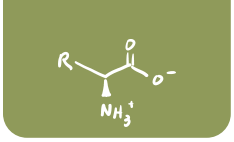
Structure	Name	Abbreviations[a]	pK_{a_1} α-CO_2H	pK_{a_2} α-NH_3^+	pK_{a_3} R group	pI
	Isoleucine[b]	I or Ile	2.4	9.7		6.1
	Phenylalanine[b]	F or Phe	1.8	9.1		5.5
	Tyrosine	Y or Tyr	2.2	9.1	10.1	5.7
	Tryptophan[b]	W or Trp	2.4	9.4		5.9
	Serine	S or Ser	2.2	9.2		5.7
	Threonine[b]	T or Thr	2.6	10.4		6.5
	Proline	P or Pro	2.0	10.6		6.3
	4-Hydroxyproline (cis and trans)	O or Hyp	1.9	9.7		6.3
	Cysteine	C or Cys	1.7	10.8	8.3	5.0

(continues on next page)

TABLE 24.1 CONTINUED

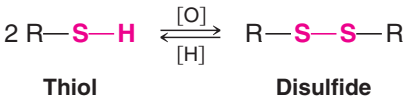
Structure	Name	Abbreviations[a]	pK_{a_1} α-CO_2H	pK_{a_2} α-NH_3^+	pK_{a_3} R group	p*I*
	Cystine	Cys-Cys	1.6 2.3	7.9 9.9		5.1
	Methionine[b]	M or Met	2.3	9.2		5.8
	Asparagine	N or Asn	2.0	8.8		5.4
	Glutamine	Q or Gln	2.2	9.1		5.7
Side Chains Containing an Acidic (Carboxyl) Group						
	Aspartic acid	D or Asp	2.1	9.8	3.9	3.0
	Glutamic acid	E or Glu	2.2	9.7	4.3	3.2
Side Chains Containing a Basic Group						
	Lysine[b]	K or Lys	2.2	9.0	10.5[c]	9.8
	Arginine	R or Arg	2.2	9.0	12.5[c]	10.8
	Histidine	H or His	1.8	9.2	6.0[c]	7.6

[a]Single-letter abbreviations are now the most commonly used form in current biochemical literature.
[b]An essential amino acid.
[c]pK_a is of protonated amine of R group.

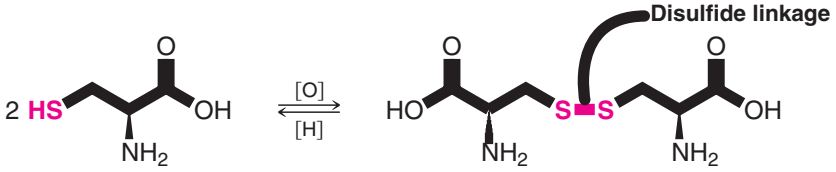


Only 20 of the 22 α-amino acids in Table 24.1 are actually used by cells when they synthesize proteins. Two amino acids are synthesized after the polyamide chain is intact. Hydroxyproline (present mainly in collagen) is synthesized by oxidation of proline, and cystine (present in most proteins) is synthesized from cysteine.

The conversion of cysteine to cystine requires additional comment. The —SH group of cysteine makes cysteine a *thiol.* One property of thiols is that they can be converted to disulfides by mild oxidizing agents. This conversion, moreover, can be reversed by mild reducing agents:



Thiol **Disulfide**



Cysteine **Cystine**

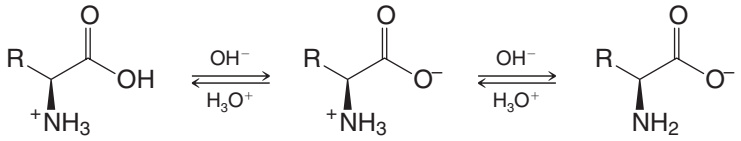
We shall see later how the **disulfide linkage** between cysteine units in a protein chain contributes to the overall structure and shape of the protein.

24.2B Essential Amino Acids

Amino acids can be synthesized by all living organisms, plants and animals. Many higher animals, however, are deficient in their ability to synthesize all of the amino acids they need for their proteins. Thus, these higher animals require certain amino acids as a part of their diet. For adult humans there are eight **essential amino acids;** these are identified in Table 24.1 by a footnote.

24.2C Amino Acids as Dipolar Ions

- Amino acids contain both a basic group (—NH_2) and an acidic group (—CO_2H).
- In the dry solid state, amino acids exist as **dipolar ions,** a form in which the carboxyl group is present as a carboxylate ion, —CO_2^-, and the amino group is present as an aminium ion, —NH_3^+ (Dipolar ions are also called **zwitterions.**)
- In aqueous solution, an equilibrium exists between the dipolar ion and the anionic and cationic forms of an amino acid.



Cationic form (predominant in strongly acidic solutions, e.g., at pH 0) **Dipolar ion** **Anionic form (predominant in strongly basic solutions, e.g., at pH 14)**

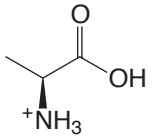
The predominant form of the amino acid present in a solution depends on the pH of the solution and on the nature of the amino acid. In strongly acidic solutions all amino acids are present primarily as cations; in strongly basic solutions they are present as anions.

- The **isoelectric point (p*I*)** is the pH at which the concentration of the dipolar ion is at its maximum and the concentrations of the anions and cations are equal.

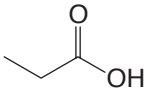
Each amino acid has a particular isoelectric point. These are given in Table 24.1. Proteins have isoelectric points as well. As we shall see later (Sections 24.13 and 24.14), this property of proteins is important for their separation and identification.

Let us consider first an amino acid with a side chain that contains neither acidic nor basic groups—an amino acid, for example, such as alanine.

If alanine is dissolved in a strongly acidic solution (e.g., pH 0), it is present in mainly a net cationic form. In this state the amino group is protonated (bears a formal +1 charge) and the carboxylic acid group is neutral (has no formal charge). As is typical of α-amino acids, the pK_a for the carboxylic acid hydrogen of alanine is considerably lower (2.3) than the pK_a of an ordinary carboxylic acid (e.g., propanoic acid, pK_a 4.89):



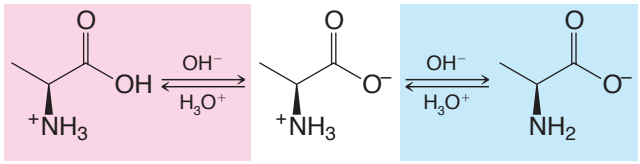
Cationic form of alanine
pK_{a_1} = 2.3



Propanoic acid
pK_a = 4.89

The reason for this enhanced acidity of the carboxyl group in an α-amino acid is the inductive effect of the neighboring aminium cation, which helps to stabilize the carboxylate anion formed when it loses a proton. Loss of a proton from the carboxyl group in a cationic α-amino acid leaves the molecule electrically neutral (in the form of a dipolar ion). This equilibrium is shown in the red-shaded portion of the equation below.

The protonated amino group of an α-amino acid is also acidic, but less so than the carboxylic acid group. The pK_a of the aminium group in alanine is 9.7. The equilibrium for loss of an aminium proton is shown in the blue-shaded portion of the equation below. The carboxylic acid proton is always lost before a proton from the aminium group in an α-amino acid.



Cationic form
(pK_{a_1} = 2.3)

Dipolar ion
(pK_{a_2} = 9.7)

Anionic form

The state of an α-amino acid at any given pH is governed by a combination of two equilibria, as shown in the above equation for alanine. The isoelectric point (pI) of an amino acid such as alanine is the average of pK_{a_1} and pK_{a_2}:

$$pI = \tfrac{1}{2}(2.3 + 9.7) = 6.0 \quad \text{(isoelectric point of alanine)}$$

When a base is added to a solution of the net cationic form of alanine (initially at pH 0, for example), the first proton removed is the carboxylic acid proton, as we have said. In the case of alanine, when a pH of 2.3 is reached, the carboxylic acid proton will have been removed from half of the molecules. This pH represents the pK_a of the alanine carboxylic acid proton, as can be demonstrated using the **Henderson–Hasselbalch equation.**

- The **Henderson–Hasselbalch equation** shows that for an acid (HA) and its conjugate base (A^-) when [HA] = [A^-], then pH = pK_a.

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

Henderson–Hasselbalch equation

Therefore, when the acid is half neutralized,

$$[HA] = [A^-],\ \log \frac{[HA]}{[A^-]} = 0, \text{ and thus } pH = pK_a$$

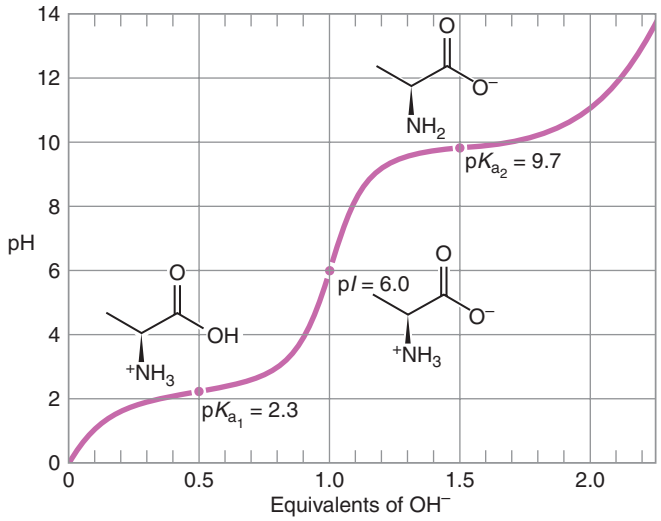


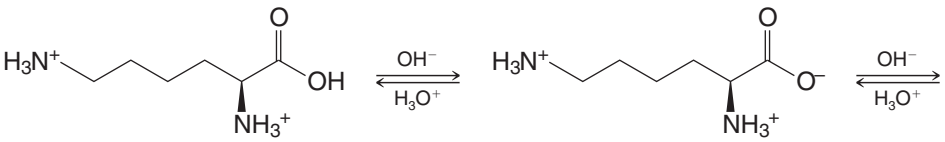
Figure 24.1 A titration curve for alanine.

As more base is added to this solution, alanine reaches its isoelectric point (p*I*), the pH at which all of alanine's carboxylic acid protons have been removed but not its aminium protons. The molecules are therefore electrically neutral (in their dipolar ion or zwitterionic form) because the carboxylate group carries a -1 charge and the aminium group a $+1$ charge. The p*I* for alanine is 6.0.

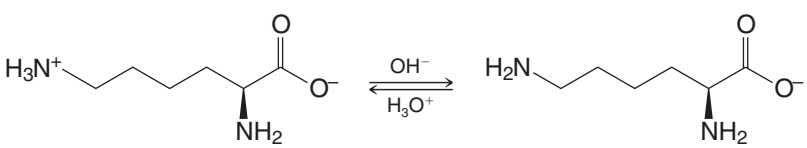
Now, as we continue to add the base, protons from the aminium ions will begin to be removed, until at pH 9.7 half of the aminium groups will have lost a proton. This pH represents the pK_a of the aminium group. Finally, as more base is added, the remaining aminium protons will be lost until all of the alanine molecules have lost their aminium protons. At this point (e.g., pH 14) the molecules carry a net anionic charge from their carboxylate group. The amino groups are now electrically neutral.

Figure 24.1 shows a titration curve for these equilibria. The graph represents the change in pH as a function of the number of molar equivalents of base. Because alanine has two protons to lose in its net cationic form, when one molar equivalent of base has been added, the molecules will have each lost one proton and they will be electrically neutral (the dipolar ion or zwitterionic form).

If an amino acid contains a side chain that has an acidic or basic group, the equilibria become more complex. Consider lysine, for example, an amino acid that has an additional —NH_2 group on its ε carbon. In strongly acidic solution, lysine is present as a dication because both amino groups are protonated. The first proton to be lost as the pH is raised is a proton of the carboxyl group (pK_{a_1} = 2.2), the next is from the α-aminium group (pK_{a_2} = 9.0), and the last is from the ε-aminium group (pK_{a_3} = 10.5):



Dicationic form of lysine (pK_{a_1} = 2.2) | **Monocationic form (pK_{a_2} = 9.0)**



Dipolar ion (pK_{a_3} = 10.5) | **Anionic form**

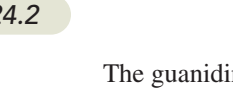
The isoelectric point of lysine is the average of pK_{a_2} the monocation) and pK_{a_3} (the dipolar ion).

$$pI = \tfrac{1}{2}(9.0 + 10.5) = 9.8 \quad \text{(isoelectric point of lysine)}$$

Review Problem 24.1

What form of glutamic acid would you expect to predominate in **(a)** strongly acidic solution, **(b)** strongly basic solution, and **(c)** at its isoelectric point (p*I* 3.2)? **(d)** The isoelectric point of glutamine (p*I* 5.7) is considerably higher than that of glutamic acid. Explain.

Review Problem 24.2

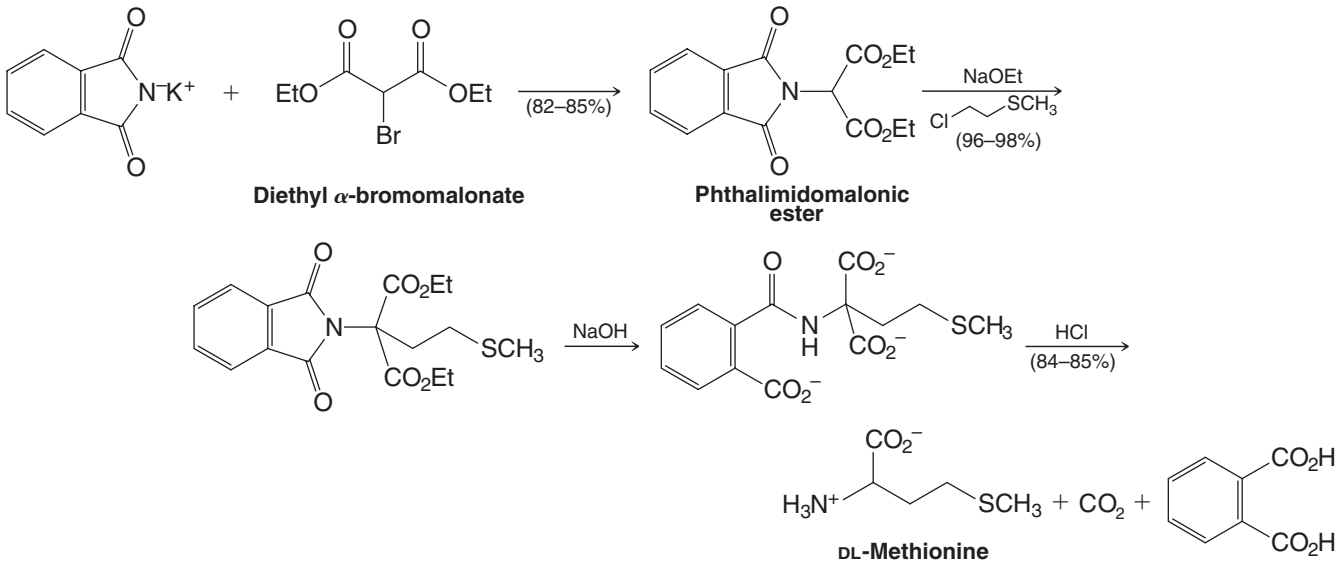
The guanidino group  of arginine is one of the most strongly basic of all organic groups. Explain.

24.3 *Synthesis of α-Amino Acids*

A variety of methods have been developed for the synthesis of α-amino acids. Here we describe two methods that are based on reactions we have studied before. In "The Chemistry of . . . Asymmetric Syntheses of Amino Acids" (*WileyPLUS*) we show methods to prepare α-amino acids in optically active form. Asymmetric synthesis is an important goal in α-amino acid synthesis due to the biological activity of the natural enantiomeric forms of α-amino acids, and due to the commercial relevance of products made by these routes.

24.3A From Potassium Phthalimide

This method, a modification of the Gabriel synthesis of amines (Section 20.4A), uses potassium phthalimide and diethyl α-bromomalonate to prepare an *imido* malonic ester. The example shown is a synthesis of methionine:

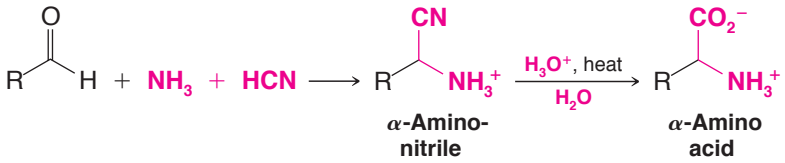


Review Problem 24.3

Starting with diethyl α-bromomalonate and potassium phthalimide and using any other necessary reagents, show how you might synthesize **(a)** DL-leucine, **(b)** DL-alanine, and **(c)** DL-phenylalanine.

24.3B The Strecker Synthesis

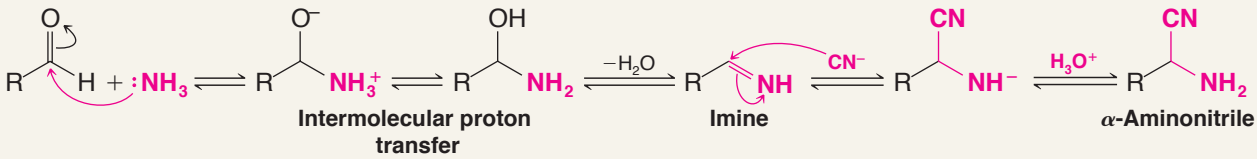
Treating an aldehyde with ammonia and hydrogen cyanide produces an α-aminonitrile. Hydrolysis of the nitrile group (Section 17.3) of the α-aminonitrile converts the latter to an α-amino acid. This synthesis is called the Strecker synthesis:



The first step of this synthesis probably involves the initial formation of an imine from the aldehyde and ammonia followed by the addition of hydrogen cyanide.

A MECHANISM FOR THE REACTION

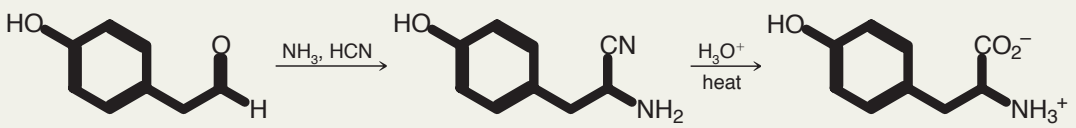
Formation of an α-Aminonitrile during the Strecker Synthesis



Solved Problem 24.1

Outline a Strecker synthesis of DL-tyrosine.

ANSWER



DL-Tyrosine

Review Problem 24.4

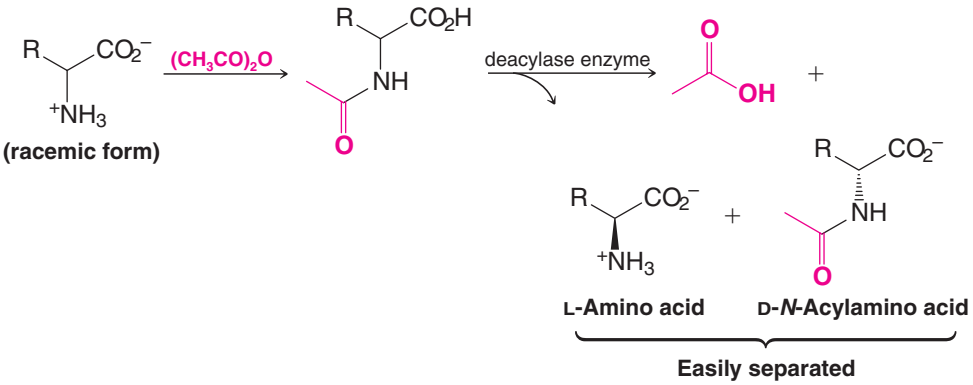
(a) Outline a Strecker synthesis of DL-phenylalanine. **(b)** DL-Methionine can also be synthesized by a Strecker synthesis. The required starting aldehyde can be prepared from acrolein ($CH_2{=}CHCHO$) and methanethiol (CH_3SH). Outline all steps in this synthesis of DL-methionine.

24.3C Resolution of DL-Amino Acids

With the exception of glycine, which has no chirality center, the amino acids that are produced by the methods we have outlined are all produced as racemic forms. To obtain the naturally occurring L-amino acid, we must, of course, resolve the racemic form. This can be done in a variety of ways, including the methods outlined in Section 20.3F.

One especially interesting method for resolving amino acids is based on the use of enzymes called *deacylases*. These enzymes catalyze the hydrolysis of *N-acylamino acids* in living organisms. Since the active site of the enzyme is chiral, it hydrolyzes only

N-acylamino acids of the L configuration. When it is exposed to a racemic mixture of *N*-acylamino acids, only the derivative of the L-amino acid is affected and the products, as a result, are separated easily:

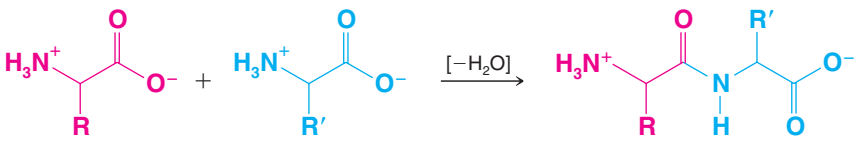


24.4 Polypeptides and Proteins

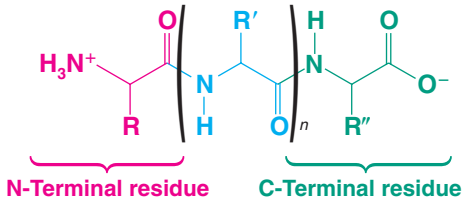
Amino acids are polymerized in living systems by enzymes that form amide linkages from the amino group of one amino acid to the carboxyl group of another.

- A molecule formed by joining amino acids together is called a **peptide,** and the amide linkages in them are called **peptide bonds** or **peptide linkages.** Each amino acid in the peptide is called an **amino acid residue.**

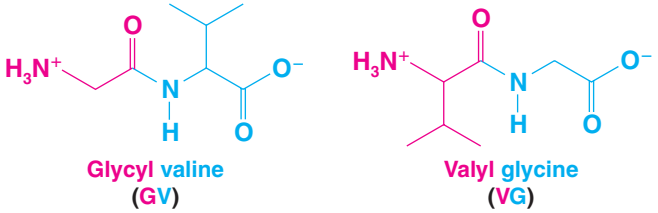
Peptides that contain 2, 3, a few (3–10), or many amino acids are called **dipeptides, tripeptides, oligopeptides,** and **polypeptides,** respectively. **Proteins** are polypeptides consisting of one or more polypeptide chains.



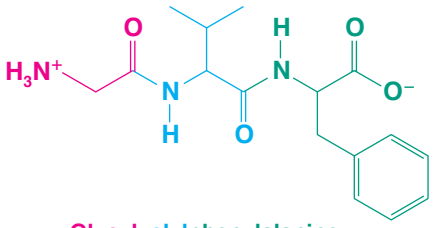
Polypeptides are **linear polymers.** One end of a polypeptide chain terminates in an amino acid residue that has a free $—NH_3^+$ group; the other terminates in an amino acid residue with a free $—CO_2^-$ group. These two groups are called the **N-terminal** and the **C-terminal residues,** respectively:



- By convention, we write peptide and protein structures with the N-terminal amino acid residue on the left and the C-terminal residue on the right:



The tripeptide glycylvalylphenylalanine has the following structural formula:



Glycylvalylphenylalanine
(GVF)

It becomes a significant task to write a full structural formula for a polypeptide chain that contains any more than a few amino acid residues. In this situation, use of the one-letter abbreviations (Table 24.1) is the norm for showing the sequence of amino acids. Very short peptide sequences are sometimes still represented with the three-letter abbreviations (Table 24.1).

24.4A Hydrolysis

When a protein or polypeptide is refluxed with 6*M* hydrochloric acid for 24 h, hydrolysis of all the amide linkages usually takes place, liberating its constitutent amino acids as a mixture. Chromatographic separation and quantitative analysis of the resulting mixture can then be used to determine which amino acids composed the intact polypeptide and their relative amounts.

One chromatographic method for separation of a mixture of amino acids is based on the use of *cation-exchange resins* (Fig. 24.2), which are insoluble polymers containing sulfonate groups. If an acidic solution containing a mixture of amino acids is passed through a column packed with a cation-exchange resin, the amino acids will be adsorbed by the resin because of attractive forces between the negatively charged sulfonate groups and the positively charged amino acids. The strength of the adsorption varies with the basicity of the individual amino acids; those that are most basic are held most strongly. If the column is then washed with a buffered solution at a given pH, the individual amino acids move down the column at different rates and ultimately become separated. In an automated version of this analysis developed at Rockefeller University in 1950, the eluate is allowed to mix with **ninhydrin,** a reagent that reacts with most amino acids to give a derivative with an intense purple color (λ_{max} 570 nm). The amino acid analyzer is designed so that it can measure the absorbance of the eluate (at 570 nm) continuously and record this absorbance as a function of the volume of the effluent.

A typical graph obtained from an automatic amino acid analyzer is shown in Fig. 24.3. When the procedure is standardized, the positions of the peaks are characteristic of the individual amino acids, and the areas under the peaks correspond to their relative amounts.

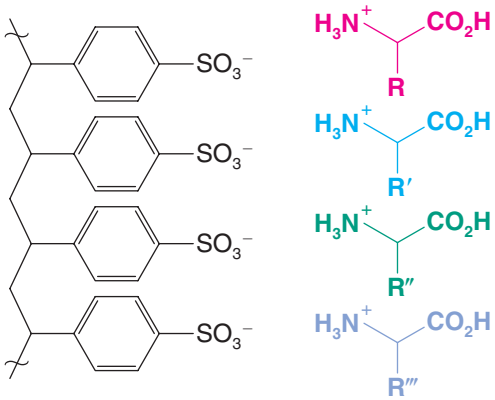


Figure 24.2 A section of a cation-exchange resin with adsorbed amino acids.

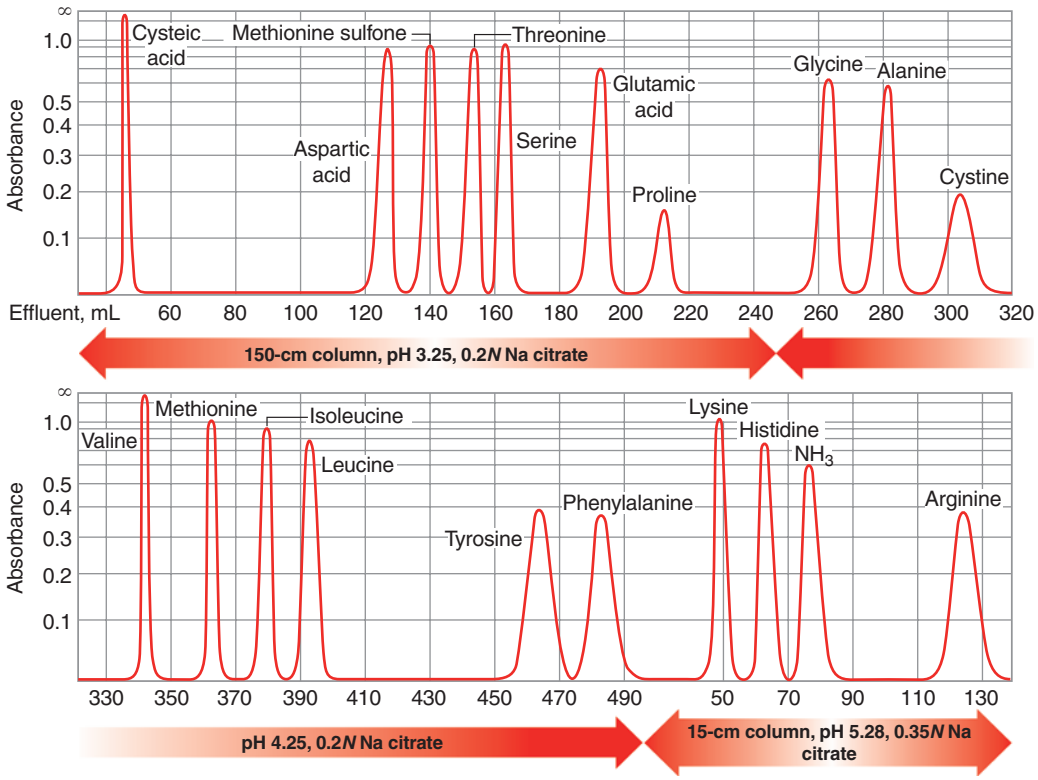
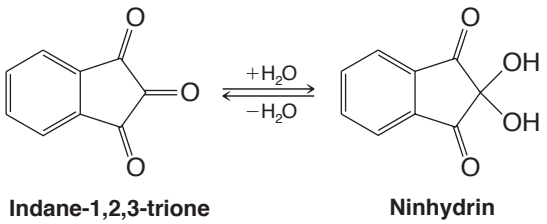
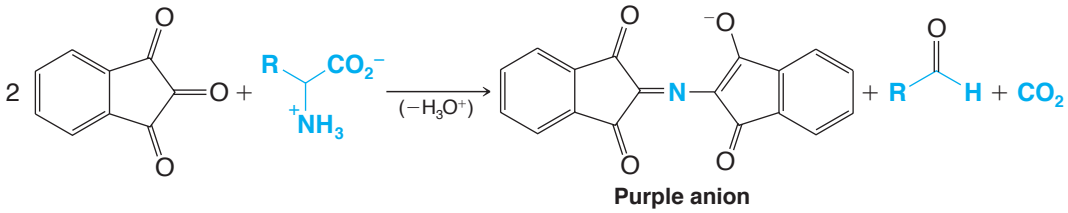


Figure 24.3 Typical result given by an automatic amino acid analyzer. (Adapted with permission from Spackman, D. H., Stein, W. H., and Moore, S., *Analytical Chemistry, 30(7)*, pp. 1190–1206, Figure 2, 1958. Copyright 1958 American Chemical Society.)

Ninhydrin is the hydrate of indane-1,2,3-trione. With the exception of proline and hydroxyproline, all of the α-amino acids found in proteins react with ninhydrin to give the same intensely colored purple anion (λ_{max} 570 nm). We shall not go into the mechanism here, but notice that the only portion of the anion that is derived from the α-amino acid is the nitrogen:



Indane-1,2,3-trione **Ninhydrin**



Purple anion

Proline and hydroxyproline do not react with ninhydrin in the same way because their α-amino groups are secondary amines and part of a five-membered ring.

Analysis of amino acid mixtures can also be done very easily using high-performance liquid chromatography (HPLC), and this is now the most common method. A cation-exchange resin is used for the column packing in some HPLC analyses (see Section 24.14),

while other analyses require hydrophobic (reversed-phase) column materials. Identification of amino acids separated by HPLC can be done by comparison with retention times of standard samples. Instruments that combine HPLC with mass spectrometry make direct identification possible (see Section 24.5E).

24.5 Primary Structure of Polypeptides and Proteins

The sequence of amino acid residues in a polypeptide or protein is called its **primary structure.** A simple peptide composed of three amino acids (a tripeptide) can have 6 different amino acid sequences; a tetrapeptide can have as many as 24 different sequences. For a protein composed of 20 different amino acids in a single chain of 100 residues, there are $2^{100} = 1.27 \times 10^{130}$ possible peptide sequences, a number much greater than the number of atoms estimated to be in the universe (9×10^{78})! Clearly, one of the most important things to determine about a protein is the sequence of its amino acids. Fortunately, there are a variety of methods available to determine the sequence of amino acids in a polypeptide. We shall begin with **terminal residue analysis** techniques used to identify the N- and C-terminal amino acids.

24.5A Edman Degradation

The most widely used procedure for identifying the N-terminal amino acid in a peptide is the **Edman degradation** method (developed by Pehr Edman of the University of Lund, Sweden). Used repetitively, the Edman degradation method can be used to sequence peptides up to about 60 residues in length. The process works so well that machines called amino acid sequencers have been developed to carry out the Edman degradation process in automated cycles.

The chemistry of the Edman degradation is based on a labeling reaction between the N-terminal amino group and phenyl isothiocyanate, $C_6H_5—N{=}C{=}S$. Phenyl isothiocyanate reacts with the N-terminal amino group to form a phenylthiocarbamyl derivative, which is then cleaved from the peptide chain by acid. The result is an unstable anilinothioazolinone (ATZ), which rearranges to a stable phenylthiohydantoin (PTH) derivative of the amino acid. In the automated process, the PTH derivative is introduced directly to a high-performance liquid chromatograph and identified by comparison of its retention time with known amino acid PTH derivatives (Fig. 24.4). The cycle is then repeated for the next N-terminal amino acid. Automated peptide sequence analyzers can

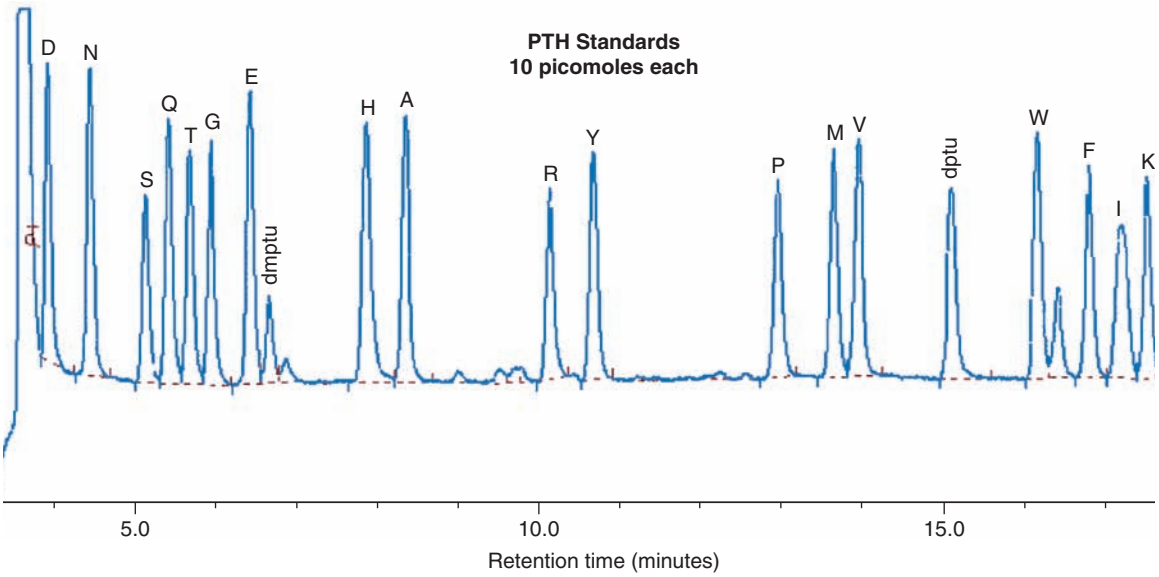
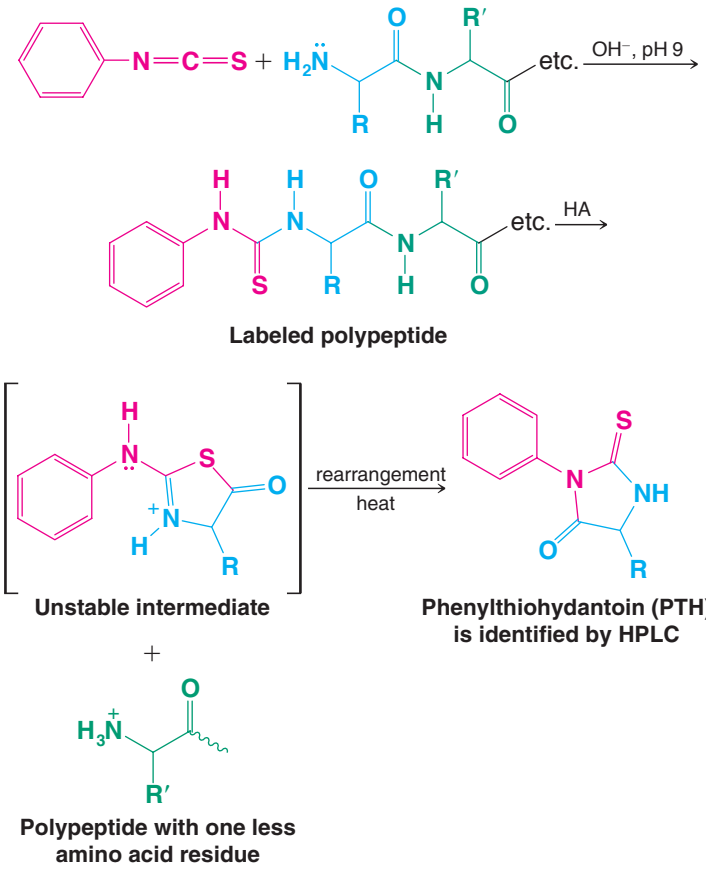


Figure 24.4 PTH amino acid standards run on a Procise instrument; see Table 24.1 for amino acid abbreviations. Peaks marked dmptu (dimethylphenylthiourea) and dptu (diphenylthiourea) represent side-reaction products of the Edman degradation. (*Courtesy of Applied Biosystems.*)

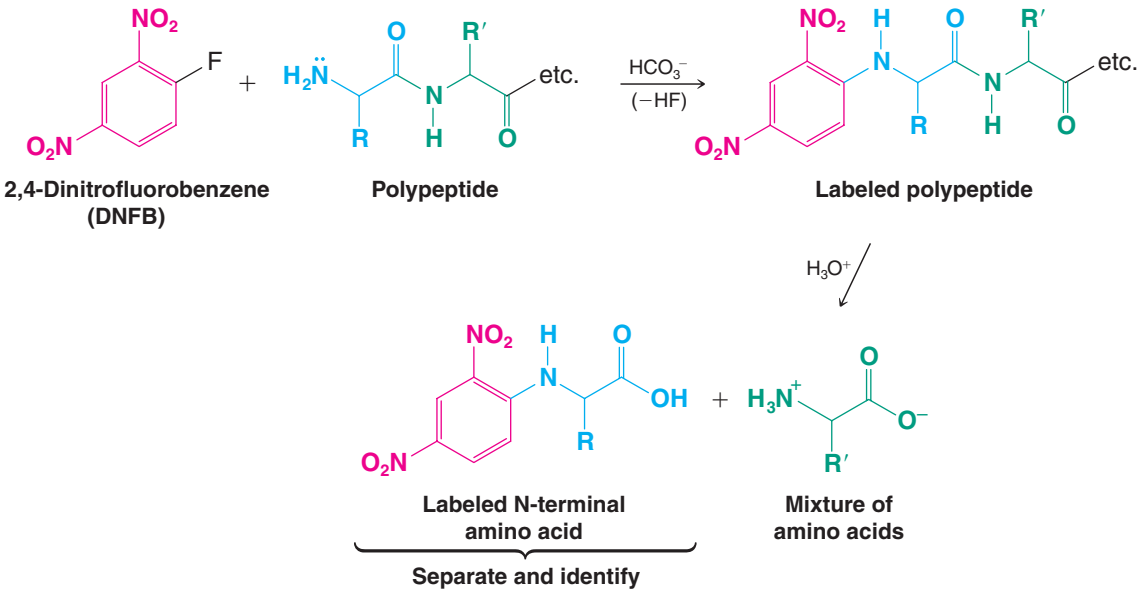
perform a single iteration of the Edman degradation in approximately 30 min using only picomole amounts of the polypeptide sample.

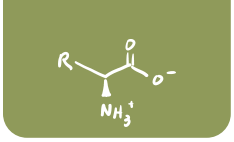


This method was introduced by Frederick Sanger of Cambridge University in 1945. Sanger made extensive use of this procedure in his determination of the amino acid sequence of insulin and won the Nobel Prize in Chemistry for the work in 1958.

24.5B Sanger N-Terminal Analysis

Another method for sequence analysis is the **Sanger N-terminal analysis,** based on the use of 2,4-dinitrofluorobenzene (DNFB). When a polypeptide is treated with DNFB in mildly basic solution, a nucleophilic aromatic substitution reaction (S_NAr, Section 21.11A) takes place involving the free amino group of the N-terminal residue. Subsequent hydrolysis of the polypeptide gives a mixture of amino acids in which the N-terminal amino acid





is labeled with a 2,4-dinitrophenyl group. After separating this amino acid from the mixture, it can be identified by comparison with known standards.

2,4-Dinitrofluorobenzene will react with any free amino group in a polypeptide, including the ε-amino group of lysine, and this fact complicates Sanger analyses. Only the N-terminal amino acid residue of a peptide will bear the 2,4-dinitrophenyl group at its α-amino group, however. Nevertheless, the Edman method of N-terminal analysis is much more widely used.

Review Problem 24.5

The electron-withdrawing property of the 2,4-dinitrophenyl group makes separation of the labeled amino acid very easy. Suggest how this is done.

24.5C C-Terminal Analysis

C-Terminal residues can be identified through the use of digestive enzymes called *carboxypeptidases.* These enzymes specifically catalyze the hydrolysis of the amide bond of the amino acid residue containing a free $-CO_2H$ group, liberating it as a free amino acid. A carboxypeptidase, however, will continue to attack the polypeptide chain that remains, successively lopping off C-terminal residues. As a consequence, it is necessary to follow the amino acids released as a function of time. The procedure can be applied to only a limited amino acid sequence for, at best, after a time the situation becomes too confused to sort out.

Review Problem 24.6

(a) Write a reaction showing how 2,4-dinitrofluorobenzene could be used to identify the N-terminal amino acid of VAG. **(b)** What products would you expect (after hydrolysis) when VKG is treated with 2,4-dinitrofluorobenzene?

Review Problem 24.7

Write the reactions involved in a sequential Edman degradation of MIR.

24.5D Complete Sequence Analysis

Sequential analysis using the Edman degradation or other methods becomes impractical with large proteins and polypeptides. Fortunately, there are techniques to cleave peptides into fragments that are of manageable size. **Partial hydrolysis** with dilute acid, for example, generates a family of peptides cleaved in random locations and with varying lengths. Sequencing these cleavage peptides and looking for points of overlap allows the sequence of the entire peptide to be pieced together.

Consider a simple example: We are given a pentapeptide known to contain valine (two residues), leucine (one residue), histidine (one residue), and phenylalanine (one residue), as determined by hydrolysis and automatic amino acid analysis. With this information we can write the "molecular formula" of the protein in the following way, using commas to indicate that the sequence is unknown:

2V, L, H, F

Then, let us assume that by using DNFB and carboxypeptidase we discover that valine and leucine are the N- and C-terminal residues, respectively. So far we know the following:

V (V, H, F) L

But the sequence of the three nonterminal amino acids is still unknown.

We then subject the pentapeptide to partial acid hydrolysis and obtain the following dipeptides. (We also get individual amino acids and larger pieces, i.e., tripeptides and tetrapeptides.)

VH + HV + VF + FL

The points of overlap of the dipeptides (i.e., H, V, and F) tell us that the original pentapeptide must have been the following:

VHVFL

Site-specific cleavage of peptide bonds is possible with enzymes and specialized reagents as well, and these methods are now more widely used than partial hydrolysis. For example, the enzyme trypsin preferentially catalyzes hydrolysis of peptide bonds on the C-terminal side of arginine and lysine. Chemical cleavage at specific sites can be done with cyanogen bromide (CNBr), which cleaves peptide bonds on the C-terminal side of methionine residues. Using these site-selective cleavage methods on separate samples of a given polypeptide results in fragments that have overlapping sequences. After sequencing the individual fragments, aligning them with each other on the basis of their overlapping sections results in a sequence for the intact protein.

24.5E Peptide Sequencing Using Mass Spectrometry and Sequence Databases

Other methods for determining the sequence of a polypeptide include mass spectrometry and comparison of partial peptide sequences with databases of known complete sequences.

Ladder Sequencing Mass spectrometry is especially powerful because sophisticated techniques allow mass analysis of proteins with very high precision. In one mass spectrometric method, called "ladder sequencing," an enzymatic digest is prepared that yields a mixture of peptide fragments that each differ in length by one amino acid residue (e.g., by use of carboxypeptidase). The digest is a family of peptides where each one is the result of cleavage of one successive residue from the chain. Mass spectrometric analysis of this mixture yields a family of peaks corresponding to the molecular weight of each peptide. Each peak in the spectrum differs from the next by the molecular weight of the amino acid that is the difference in their structures. With these data, one can ascend the ladder of peaks from the lowest weight fragment to the highest (or vice versa), "reading" the sequence of the peptide from the difference in mass between each peak. The difference in mass between each peptide fragment and the next represents the amino acid in that spot along the sequence, and hence an entire sequence can be read from the ladder of fragment masses. This technique has also been applied to the sequencing of oligonucleotides.

Tandem Mass Spectrometry (MS/MS) Random cleavage of a peptide, similar to that from partial hydrolysis with acid, can also be accomplished with mass spectrometry. An intact protein introduced into a mass spectrometer can be cleaved into smaller fragments by collision with gas molecules deliberately leaked into the mass spectrometer vacuum chamber (a technique called collision-induced dissociation, CID). These peptide fragments can be individually selected for analysis using a technique called tandem mass spectrometry (MS/MS). The mass spectra of these random fragments can be compared with mass spectra databases to determine the protein sequence.

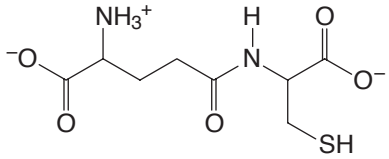
Partial Hydrolysis and Sequence Comparison In some cases it is also possible to determine the sequence of an unknown polypeptide by sequencing just a few of its amino acids and comparing this partial sequence with the database of known sequences for complete polypeptides or proteins. This procedure works if the unknown peptide turns out to be one that has been studied previously. (Studies of the expression of known proteins is one dimension of the field of proteomics, Section 24.14.) Due to the many sequence permutations that are theoretically possible and the uniqueness of a given protein's structure, a sequence of just 10–25 peptide residues is usually sufficient to generate data that match only one or a small number of known polypeptides. The partial sequence can be determined by the Edman method or by mass spectrometry. For example, the enzyme lysozyme with 129 amino acid residues (see Section 24.10) can be identified based on the sequence of just its first 15 amino acid residues. Structure determination based on comparison of sequences with computerized databases is part of the burgeoning field of bioinformatics.

An analogous approach using databases is to infer the *DNA sequence* that codes for a partial peptide sequence and compare this DNA sequence with the database of known DNA sequences. If a satisfactory match is found, the remaining sequence of the polypeptide can be read from the DNA sequence using the genetic code (see Section 25.5). In addition, the

inferred oligonucleotide sequence for the partial peptide can be synthesized chemically (see Section 25.7) and used as a probe to find the gene that codes for the protein. This technique is part of molecular biological methods used to clone and express large quantities of a protein of interest.

Review Problem 24.8

Glutathione is a tripeptide found in most living cells. Partial acid-catalyzed hydrolysis of glutathione yields two dipeptides, CG and one composed of E and C. When this second dipeptide was treated with DNFB, acid hydrolysis gave *N*-labeled glutamic acid. **(a)** On the basis of this information alone, what structures are possible for glutathione? **(b)** Synthetic experiments have shown that the second dipeptide has the following structure:



What is the structure of glutathione?

Review Problem 24.9

Give the amino acid sequence of the following polypeptides using only the data given by partial acidic hydrolysis:

(a) S, O, P, T $\xrightarrow[H_2O]{H_3O^+}$ ST + TO + PS

(b) A, R, C, V, L $\xrightarrow[H_2O]{H_3O^+}$ AC + CR + RV + LA

24.6 Examples of Polypeptide and Protein Primary Structure

- The covalent structure of a protein or polypeptide is called its **primary structure** (Fig. 24.5).

Using the techniques we described, chemists have had remarkable success in determining the primary structures of polypeptides and proteins. The compounds described in the following pages are important examples.

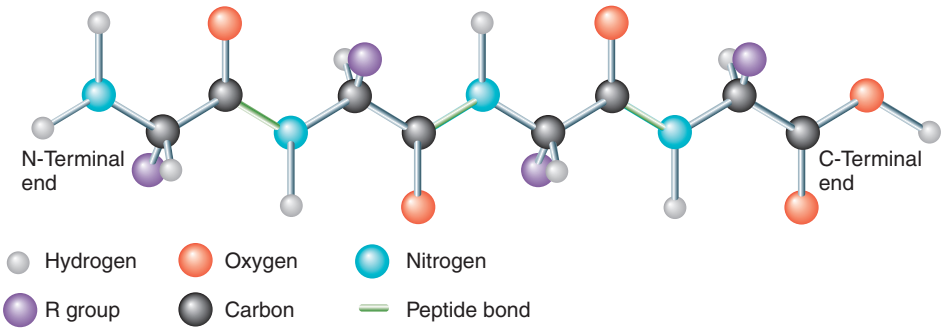


Figure 24.5 A representation of the primary structure of a tetrapeptide.

24.6A Oxytocin and Vasopressin

Oxytocin and vasopressin (Fig. 24.6) are two rather small polypeptides with strikingly similar structures (where oxytocin has leucine, vasopressin has arginine, and where oxytocin has isoleucine, vasopressin has phenylalanine). In spite of the similarity of their amino acid sequences, these two polypeptides have quite different physiological effects. Oxytocin occurs only in the female of a species and stimulates uterine contractions during childbirth. Vasopressin occurs in males and females; it causes contraction of peripheral blood vessels

Vincent du Vigneaud of Cornell Medical College synthesized oxytocin and vasopressin in 1953; he received the Nobel Prize in Chemistry in 1955.

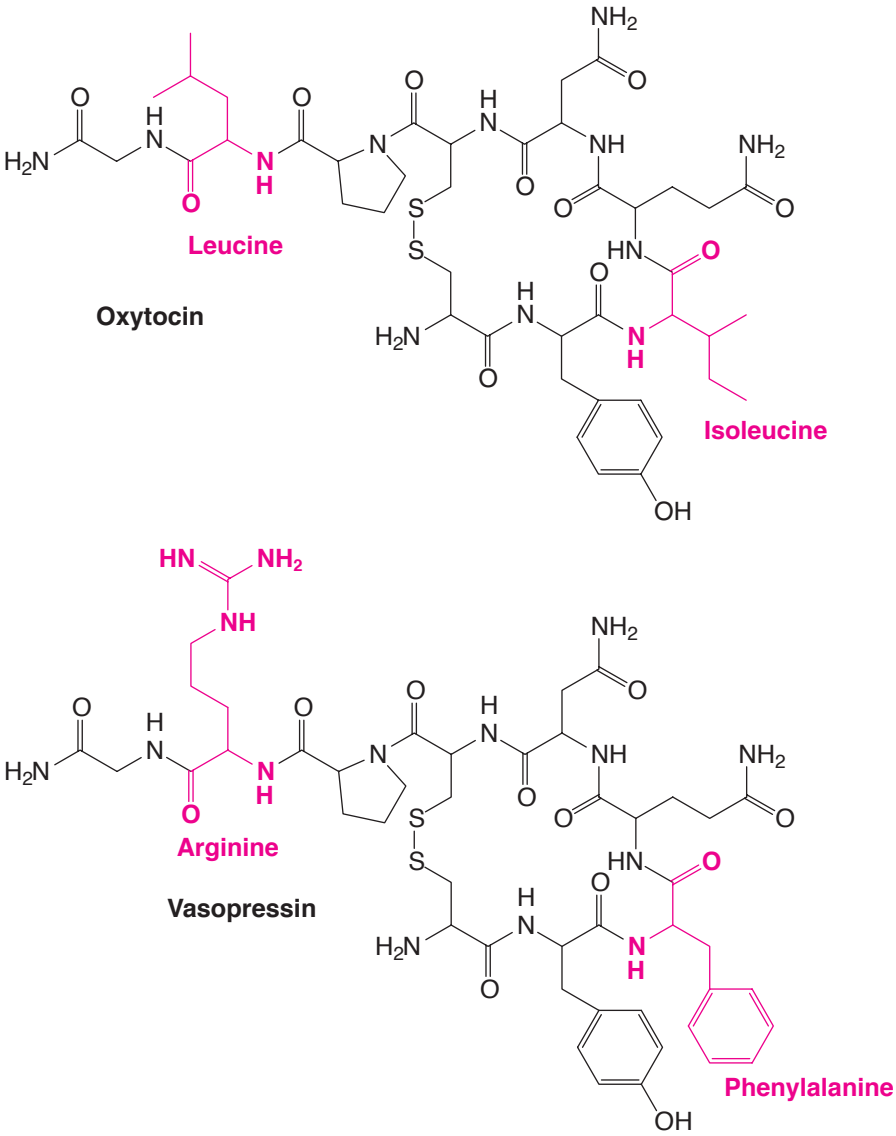


Figure 24.6 The structures of oxytocin and vasopressin. Amino acid residues that differ between them are shown in red.

and an increase in blood pressure. Its major function, however, is as an *antidiuretic;* physiologists often refer to vasopressin as an *antidiuretic hormone.*

The structures of oxytocin and vasopressin also illustrate the importance of the disulfide linkage between cysteine residues (Section 24.2A) in the overall primary structure of a polypeptide. In these two molecules this disulfide linkage leads to a cyclic structure.

Review Problem 24.10

Treating oxytocin with certain reducing agents (e.g., sodium in liquid ammonia) brings about a single chemical change that can be reversed by air oxidation. What chemical changes are involved?

24.6B Insulin

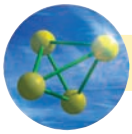
Insulin, a hormone secreted by the pancreas, regulates glucose metabolism. Insulin deficiency in humans is the major problem in diabetes mellitus.

The amino acid sequence of bovine insulin (Fig. 24.7) was determined by Sanger in 1953 after 10 years of work. Bovine insulin has a total of 51 amino acid residues in two polypeptide chains, called the A and B chains. These chains are joined by two disulfide linkages.

Figure 24.7 The amino acid sequence of bovine insulin. Lines between chains indicate disulfide linkages.

The A chain contains an additional disulfide linkage between cysteine residues at positions 6 and 11.

Human insulin differs from bovine insulin at only three amino acid residues: Threonine replaces alanine once in the A chain (residue 8) and once in the B chain (residue 30), and isoleucine replaces valine once in the A chain (residue 10). Insulins from most mammals have similar structures.

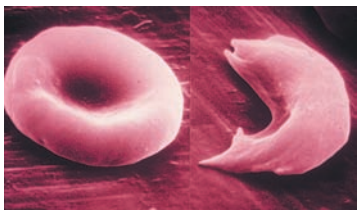


THE CHEMISTRY OF . . .

Sickle-Cell Anemia

The genetically based disease sickle-cell anemia results from a single amino acid error in the β chain of hemoglobin. In normal hemoglobin, position 6 has a glutamic acid residue, whereas in sickle-cell hemoglobin position 6 is occupied by valine.

Red blood cells (erythrocytes) containing hemoglobin with this amino acid residue error tend to become crescent shaped ("sickle") when the partial pressure of oxygen is low, as it is in venous blood. These distorted cells are more difficult for the heart to pump through small capillaries. They may even block capillaries by clumping together; at other times the red cells may even split open. Children who inherit this genetic trait from both parents suffer from a severe form of the disease and usually do not live past the age of two. Children who inherit the disease from only one parent generally have a much milder form. Sickle-cell anemia arose among the populations of central and western Africa where, ironically, it may have had a beneficial effect. People with a mild form of the disease are far less susceptible to malaria than those with normal hemoglobin. Malaria, a disease caused by an infectious microorganism, is especially prevalent in central and western Africa. Mutational changes such as those that give rise to sickle-cell anemia are very common. Approximately 150 different types of mutant hemoglobin have been detected in humans; fortunately, most are harmless.



Normal (left) and sickled (right) red blood cells viewed with a scanning electron microscope at 18,000× magnification.

24.6C Other Polypeptides and Proteins

Successful sequential analyses have now been achieved with thousands of other polypeptides and proteins, including the following:

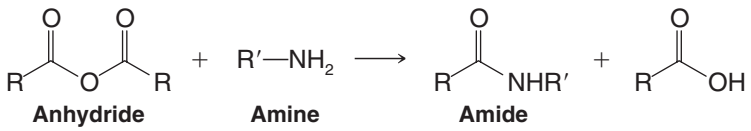
1. **Bovine ribonuclease.** This enzyme, which catalyzes the hydrolysis of ribonucleic acid (Chapter 25), has a single chain of 124 amino acid residues and four intrachain disulfide linkages.
2. **Human hemoglobin.** There are four peptide chains in this important oxygen-carrying protein. Two identical α chains have 141 residues each, and two identical β chains have 146 residues each.

3. **Bovine trypsinogen and chymotrypsinogen.** These two digestive enzyme precursors have single chains of 229 and 245 residues, respectively.
4. **Gamma globulin.** This immunoprotein has a total of 1320 amino acid residues in four chains. Two chains have 214 residues each; the other two have 446 each.
5. **p53, an anticancer protein.** The protein called p53 (the p stands for protein), consisting of 393 amino acid residues, has a variety of cellular functions, but the most important ones involve controlling the steps that lead to cell growth. It acts as a **tumor suppressor** by halting abnormal growth in normal cells, and by doing so it prevents cancer. Discovered in 1979, p53 was originally thought to be a protein synthesized by an oncogene (a gene that causes cancer). Research has shown, however, that the form of p53 originally thought to have this cancer-causing property was a mutant form of the normal protein. The unmutated (or *wild type*) p53 apparently coordinates a complex set of responses to changes in DNA that could otherwise lead to cancer. When p53 becomes mutated, it no longer provides the cell with its cancer-preventing role; it apparently does the opposite, by acting to increase abnormal growth.

   More than half of the people diagnosed with cancer each year have a mutant form of p53 in their cancers. Different forms of cancer have been shown to result from different mutations in the protein, and the list of cancer types associated with mutant p53 includes cancers of most of the body parts: brain, breast, bladder, cervix, colon, liver, lung, ovary, pancreas, prostate, skin, stomach, and so on.
6. ***Ras* proteins.** *Ras* proteins are modified proteins associated with cell growth and the cell's response to insulin. They belong to a class of proteins called prenylated proteins, in which lipid groups derived from isoprenoid biosynthesis (Special Topic E, *WileyPLUS*) are appended as thioethers to C-terminal cysteine residues. Certain mutated forms of *ras* proteins cause oncogenic changes in various eukaryotic cell types. One effect of prenylation and other lipid modifications of proteins is to anchor these proteins to cellular membranes. Prenylation may also assist with molecular recognition of prenylated proteins by other proteins.*

24.7 Polypeptide and Protein Synthesis

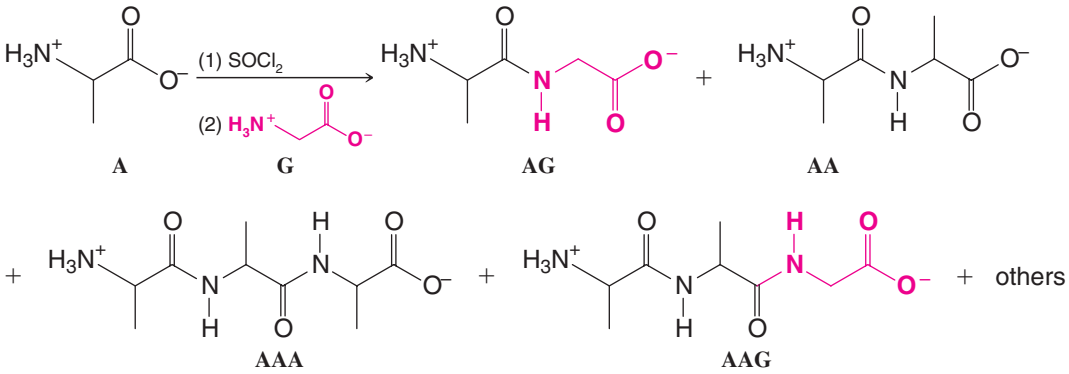
We saw in Chapter 17 that the synthesis of an amide linkage is a relatively simple one. We must first "activate" the carboxyl group of an acid by converting it to an anhydride or acid chloride and then allow it to react with an amine:



Anhydride **Amine** **Amide**

The problem becomes somewhat more complicated, however, when both the acid group and the amino group are present in the same molecule, as they are in an amino acid, and especially when our goal is the synthesis of a naturally occurring polyamide where the sequence of different amino acids is all important. Let us consider, as an example, the synthesis of the simple dipeptide alanylglycine, AG. We might first activate the carboxyl group of alanine by converting it to an acid chloride, and then we might allow it to react with glycine. Unfortunately, however, we cannot prevent alanyl chloride from reacting with itself. So our reaction would yield not only AG but also AA. It could also lead to AAA and AAG, and so on. The yield of our desired product would be low, and we would also have a difficult problem separating the dipeptides, tripeptides, and higher peptides.

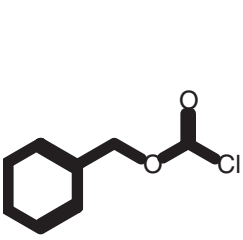
*See Gelb, M. H., "Modification of Proteins by Prenyl Groups," in *Principles of Medical Biology,* Vol. 4 (Bittar, E. E., and Bittar, N., eds.), JAI Press: Greenwich, CT, 1995; Chapter 14, pp. 323–333.



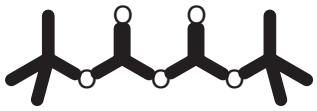
24.7A Protecting Groups

The solution to this problem is to "protect" the amino group of the first amino acid before we activate it and allow it to react with the second. By protecting the amino group, we mean that we must convert it to some other group of low nucleophilicity—*one that will not react with a reactive acyl derivative.* The **protecting group** must be carefully chosen because after we have synthesized the amide linkage between the first amino acid and the second, we will want to be able to remove the protecting group without disturbing the new amide bond.

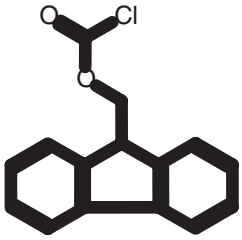
A number of reagents have been developed to meet these requirements. Three that are often used are benzyl chloroformate, di-*tert*-butyl dicarbonate (sometimes abbreviated Boc_2O, where Boc stands for *tert*-butyloxycarbonyl), and 9-fluorenylmethyl chloroformate:



Benzyl chloroformate



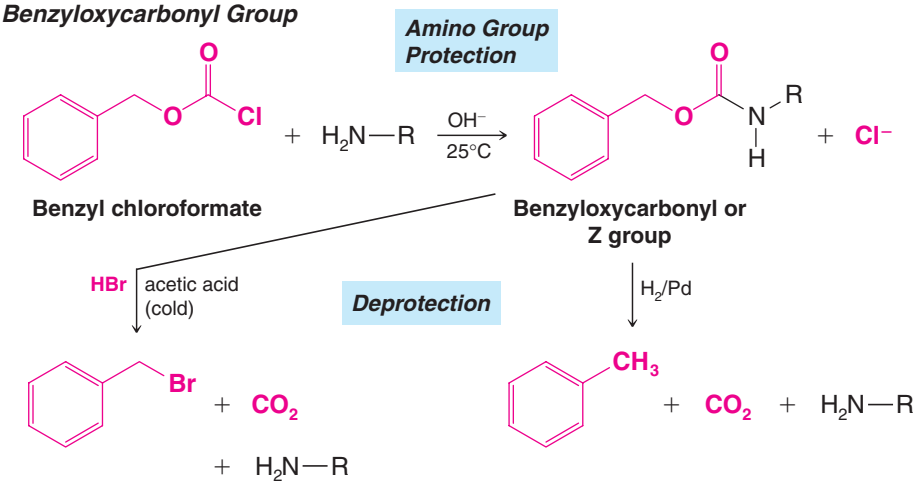
Di-*tert*-butyl dicarbonate (Boc_2O)

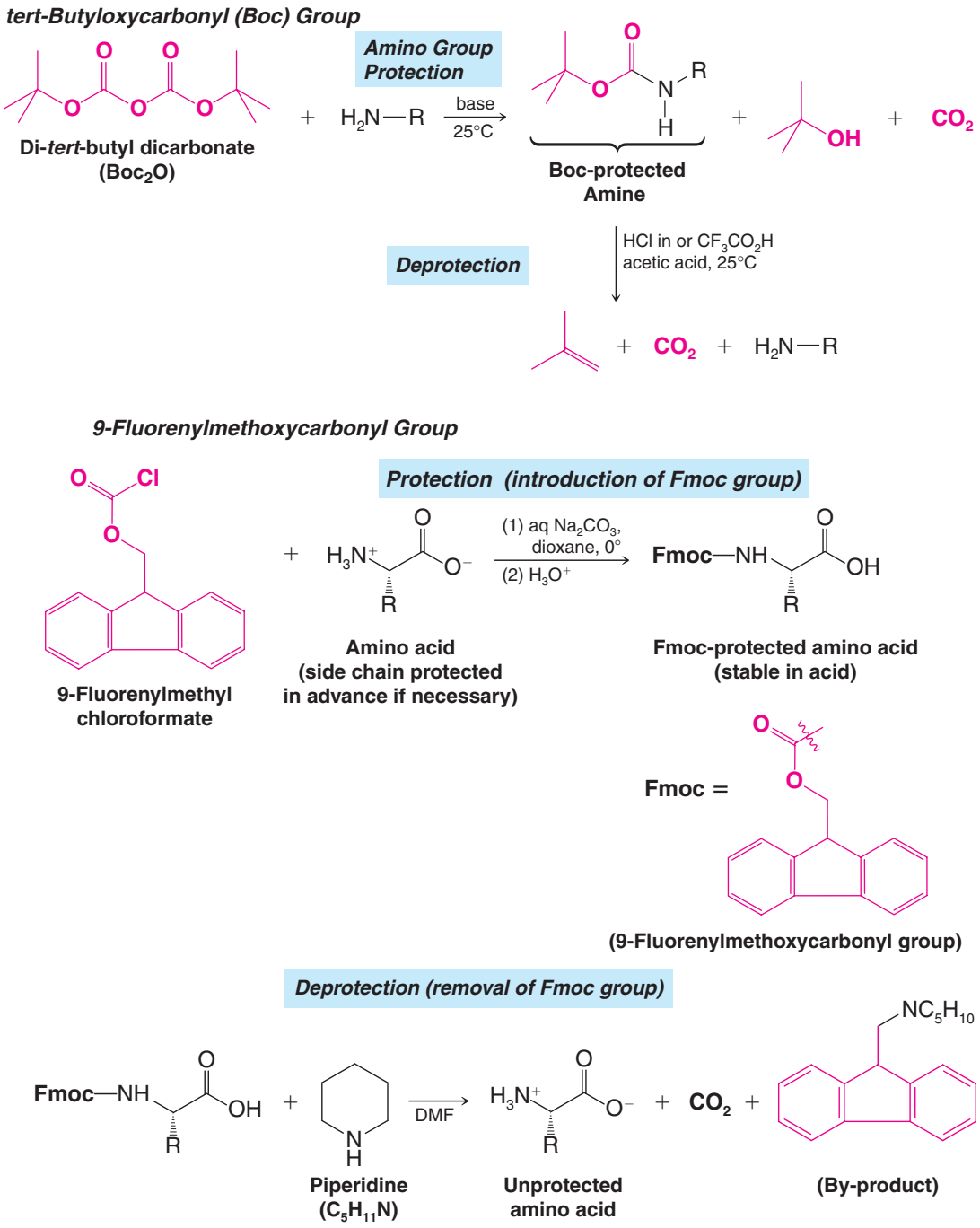


9-Fluorenylmethyl chloroformate

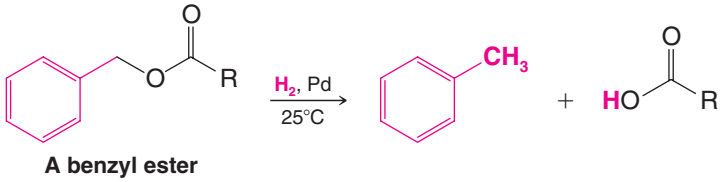
All three reagents react with the amine to block it from further acylation. These derivations, however, are types that allow removal of the protecting group under conditions that do not affect peptide bonds. The benzyloxycarbonyl group (abbreviated Z) can be removed with catalytic hydrogenation or cold HBr in acetic acid. The *tert*-butyloxycarbonyl group can be removed with trifluoroacetic acid (CF_3CO_2H) in acetic acid. The 9-fluorenylmethoxycarbonyl (Fmoc) group is stable under acid conditions but can be removed under mild basic conditions using piperidine (a secondary amine).

Benzyloxycarbonyl Group





The easy removal of the Z and Boc groups in acidic media results from the exceptional stability of the carbocations that are formed initially. The benzyloxycarbonyl group gives a benzyl carbocation; the *tert*-butyloxycarbonyl group yields, initially, a *tert*-butyl cation. Removal of the benzyloxycarbonyl group with hydrogen and a catalyst depends on the fact that benzyl–oxygen bonds are weak and subject to hydrogenolysis at low temperatures, resulting in methylbenzene (toluene) as one product:

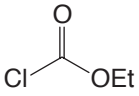


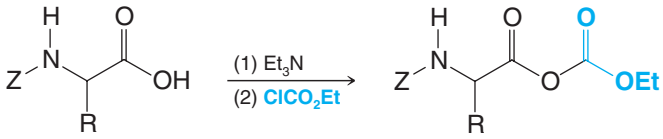
Review Problem 24.11

What classes of reactions are involved in the cleavage of the Fmoc group with piperidine, leading to the unprotected amino acid and the fluorene by-product? Write mechanisms for these reactions.

24.7B Activation of the Carboxyl Group

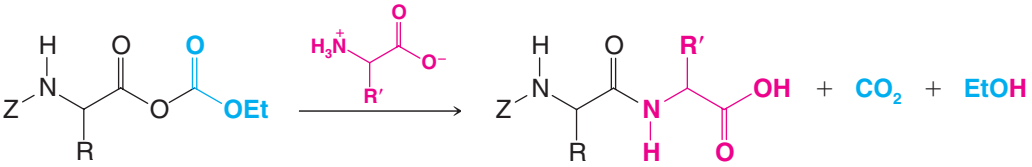
Perhaps the most obvious way to activate a carboxyl group is to convert it to an acyl chloride. This method was used in early peptide syntheses, but acyl chlorides are actually more reactive than necessary. As a result, their use leads to complicating side reactions. A much better method is to convert the carboxyl group of the "protected" amino acid to a

mixed anhydride using ethyl chloroformate, 



"Mixed anhydride"

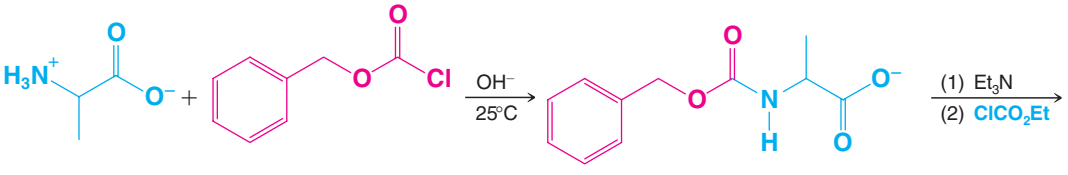
The mixed anhydride can then be used to acylate another amino acid and form a peptide linkage:



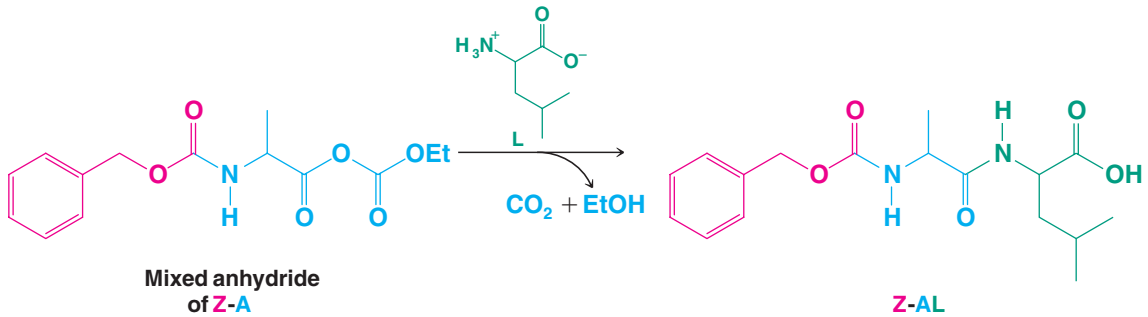
Diisopropylcarbodiimide and dicyclohexylcarbodiimide (Section 17.8E) can also be used to activate the carboxyl group of an amino acid. In Section 24.7D we shall see how diisopropylcarbodiimide is used in an automated peptide synthesis.

24.7C Peptide Synthesis

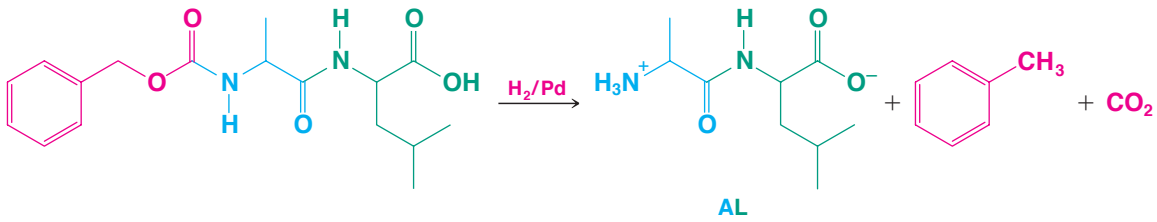
Let us examine now how we might use these reagents in the preparation of the simple dipeptide AL. The principles involved here can, of course, be extended to the synthesis of much longer polypeptide chains.



A **Benzyl chloroformate** **Z-A**



Mixed anhydride of Z-A **Z-AL**



Review Problem 24.12 Show all steps in the synthesis of GVA using the *tert*-butyloxycarbonyl (Boc) group as a protecting group.

Review Problem 24.13 The synthesis of a polypeptide containing lysine requires the protection of both amino groups. **(a)** Show how you might do this in a synthesis of the dipeptide KI using the benzyloxycarbonyl group as a protecting group. **(b)** The benzyloxycarbonyl group can also be used to protect the guanidino group, —NHC(=NH)—NH$_2$, of arginine. Show a synthesis of the dipeptide RA.

Review Problem 24.14 The terminal carboxyl groups of glutamic acid and aspartic acid are often protected through their conversion to benzyl esters. What mild method could be used for removal of this protecting group?

24.7D Automated Peptide Synthesis

R. B. Merrifield, 1984

The methods that we have described thus far have been used to synthesize a number of polypeptides, including ones as large as insulin. They are extremely time-consuming and tedious, however. One must isolate the peptide and purify it by lengthy means at almost every stage. Furthermore, significant loss of the peptide can occur with each isolation and purification stage. The development of a procedure by R. B. Merrifield (Rockefeller University, dec. 2005) for automating this process was therefore a breakthrough in peptide synthesis. Merrifield's method, for which he received the 1984 Nobel Prize in Chemistry, is called **solid-phase peptide synthesis (SPPS),** and it hinges on synthesis of the peptide residue by residue while one end of the peptide remains attached to an insoluble plastic bead. Protecting groups and other reagents are still necessary, but because the peptide being synthesized is anchored to a solid support, by-products, excess reagents, and solvents can simply be rinsed away between each synthetic step without need for intermediate purification. After the very last step the polypeptide is cleaved from the polymer support and subjected to a final purification by HPLC. The method works so well that it was developed into an automated process.

Solid-phase peptide synthesis (Fig. 24.8) begins with attachment of the first amino acid by its carboxyl group to the polymer bead, usually with a linker or spacer molecule in between. Each new amino acid is then added by formation of an amide bond between the N-terminal amino group of the peptide growing on the solid support and the new amino acid's carboxyl group. Diisopropylcarbodiimide (similar in reactivity to DCC, Section 17.8E) is used as the amide bond-forming reagent. To prevent undesired reactions as each new residue is coupled, a protecting group is used to block the amino group of the residue being added. Once the new amino acid has been coupled to the growing peptide and before the next residue is added, the protecting group on the new N-terminus is removed, making the peptide ready to begin the next cycle of amide bond formation.

Although Merrifield's initial method for solid-phase peptide synthesis used the Boc group to protect the α-amino group of residues being coupled to the growing peptide, several

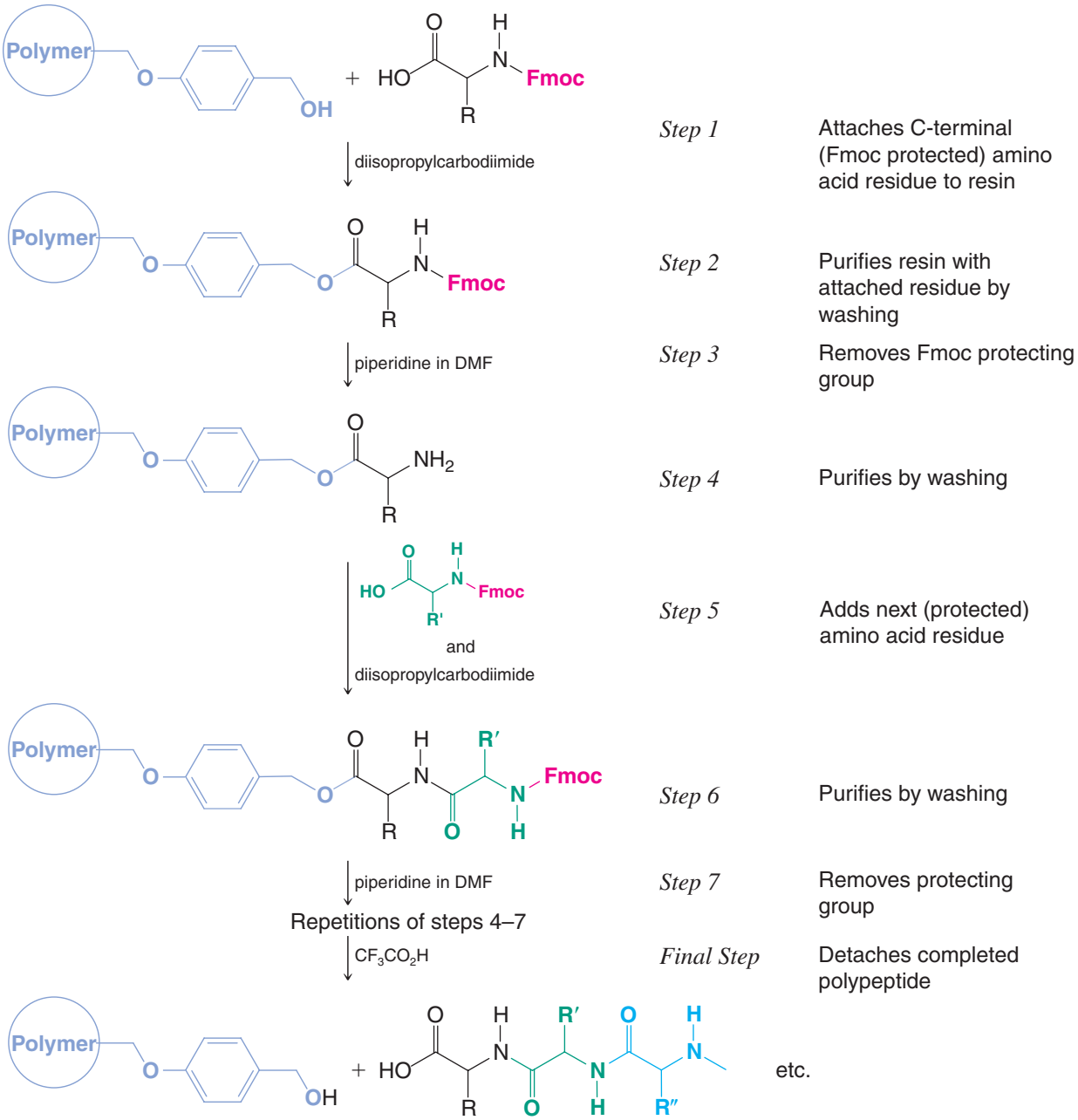


Figure 24.8 A method for automated solid-phase peptide synthesis.

advantages of the Fmoc group have since made it the group of choice. The reasons have mainly to do with excellent selectivity for removing the Fmoc group in the presence of other protecting groups used to block reactive side chains along the growing peptide and the ability to monitor the progress of the solid-phase synthesis by spectrophotometry as the Fmoc group is released in each cycle.

Let us discuss the choice of protecting groups further. As noted (Section 24.7A), *basic conditions* (piperidine in DMF) are used to remove the Fmoc group. On the other hand, protecting groups for the side chains of the peptide residues are generally blocked with *acid-labile* moieties. The base-labile Fmoc groups and acid-labile side-chain protecting groups are said to be **orthogonal protecting groups** because one set of protecting groups is stable under conditions for removal of the other, and vice versa. Another advantage of Fmoc

as compared to Boc groups for protecting the α-amino group of each new residue is that repetitive application of the acidic conditions to remove Boc groups from each new residue slowly sabotages the synthesis by prematurely cleaving some peptide molecules from the solid support and deprotecting some of the side chains. The basic conditions for Fmoc removal avoid these problematic side reactions.

- The great advantage of solid-phase peptide synthesis is that purification of the peptide at each stage involves simply rinsing the beads of the solid support to wash away excess reagent, by-products, and solvents.
- Furthermore, having the peptide attached to a tangible solid during the synthesis allows all of the steps in the synthesis to be carried out by a machine in repeated cycles.

Automated peptide synthesizers are available that can complete one cycle in 40 min and carry out 45 cycles of unattended operation. Though not as efficient as protein synthesis in the body, where enzymes directed by DNA can catalyze assembly of a protein with 150 amino acids in about 1 min, automated peptide synthesis is a far cry from the tedious process of manually synthesizing a peptide step after step. A hallmark example of automated peptide synthesis was the synthesis of ribonuclease, a protein with 124 amino acid residues. The synthesis involved 369 chemical reactions and 11,930 automated steps—all carried out without isolating an intermediate. The synthetic ribonuclease not only had the same physical characteristics as the natural enzyme, it possessed the identical biological activity as well. The overall yield was 17%, which means that the average yield of each individual step was greater than 99%.

Review Problem 24.15 One type of insoluble support used for SPPS is polymer-bound 4-benzyloxybenzyl alcohol, also known as "Wang resin," shown in Fig. 24.8. The 4-benzyloxybenzyl alcohol moiety serves as a linker between the resin backbone and the peptide. After purification, the completed polypeptide can be detached from the resin using trifluoroacetic acid under conditions that are mild enough not to affect the amide linkages. What structural features of the linker make this possible?

Review Problem 24.16 Outline the steps in the synthesis of the tripeptide KFA using the SPPS procedure.

24.8 Secondary, Tertiary, and Quaternary Structures of Proteins

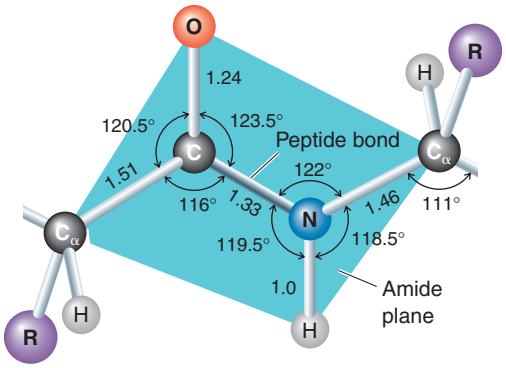
We have seen how amide and disulfide linkages constitute the covalent or *primary structure* of proteins. Of equal importance in understanding how proteins function is knowledge of the way in which the peptide chains are arranged in three dimensions. The secondary and tertiary structures of proteins are involved here.

24.8A Secondary Structure

- The **secondary structure** of a protein is defined by the local conformation of its polypeptide backbone.
- Secondary structures are specified in terms of regular folding patterns called **α helices, β-pleated sheets,** and **coil** or **loop conformations.**

To understand how these interactions occur, let us look first at what X-ray crystallographic analysis has revealed about the geometry at the peptide bond itself.

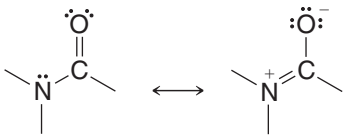
- Peptide bonds tend to assume a geometry such that six atoms of the amide linkage are coplanar (Fig. 24.9).



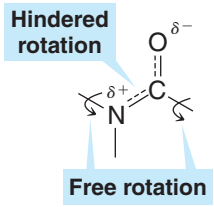
trans-Peptide group

Figure 24.9 The geometry and bond lengths (in angstroms, Å) of the peptide linkage. The six enclosed atoms tend to be coplanar and assume a "transoid" arrangement.(Reprinted with permission of John Wiley & Sons, Inc., from Voet, D. and Voet, J. G., *Biochemistry*, Second Edition. © 1995 Voet, D. and Voet, J. G.)

The carbon–nitrogen bond of the amide linkage is unusually short, indicating that resonance contributions of the type shown here are important:



Contributing resonance structures for an amide.

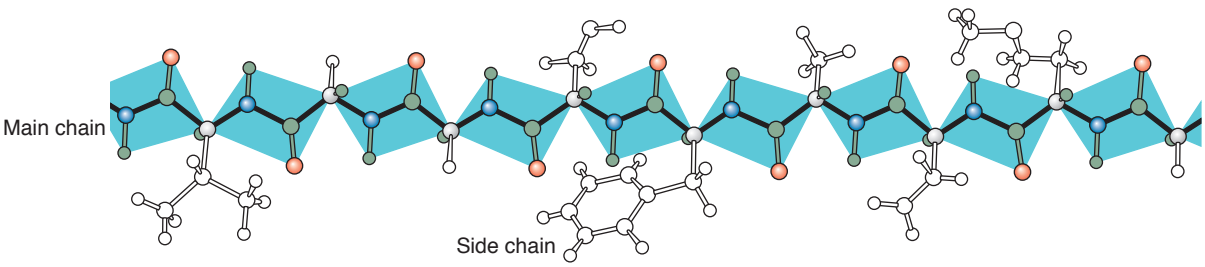


Hybrid resonance structure for an amide.

Two American scientists, Linus Pauling and Robert B. Corey, were pioneers in the X-ray analysis of proteins. Beginning in 1939, Pauling and Corey initiated a long series of studies of the conformations of peptide chains. At first, they used crystals of single amino acids, then dipeptides and tripeptides, and so on. Moving on to larger and larger molecules and using the precisely constructed molecular models, they were able to understand the secondary structures of proteins for the first time. Pauling won the 1954 Nobel Prize in Chemistry and the 1962 Nobel Peace Prize.

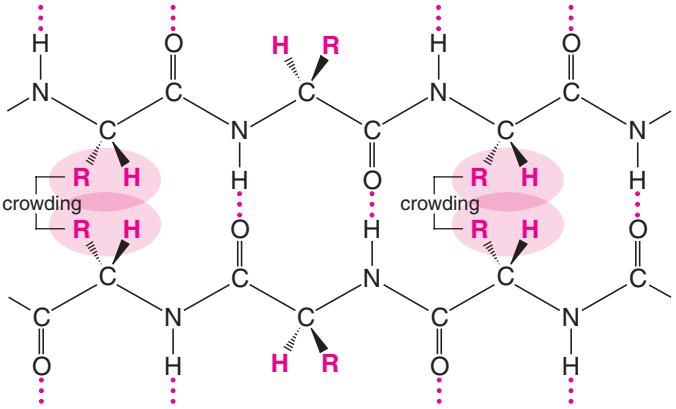
- The amide carbon–nitrogen bond, consequently, has considerable double-bond character (~40%), and rotations of groups about this bond are severely hindered.
- Rotations of groups attached to the amide nitrogen and the carbonyl carbon are relatively free, however, and these rotations allow peptide chains to form different conformations.

A transoid arrangement of groups around the relatively rigid amide bond would cause the side-chain R groups to alternate from side to side of a single fully extended peptide chain:



Calculations show that such a polypeptide chain would have a repeat distance (i.e., distance between alternating units) of 7.2 Å.

Fully extended polypeptide chains could hypothetically form a flat-sheet structure, with each alternating amino acid in each chain forming two hydrogen bonds with an amino acid in the adjacent chain:



Hypothetical flat-sheet structure
(not formed because of steric hindrance)

However, this structure does not exist in naturally occurring proteins because of the crowding that would exist between R groups. If such a structure did exist, it would have the same repeat distance as the fully extended peptide chain, that is, 7.2 Å.

- Many proteins incorporate a ***β*-pleated sheet** or ***β* configuration** (Fig. 24.10).

In a β-pleated sheet structure, slight bond rotations from one planar amide group to the next relieve the steric strain from small- and medium-sized R groups. This allows amide groups on adjacent polypeptide segments to form hydrogen bonds between the chains (see Fig. 24.10). The β-pleated sheet structure has a repeat distance of 7.0 Å between amide groups in a chain. The predominant secondary structure in silk fibroin (48% glycine and 38% serine and alanine residues) is the β-pleated sheet.

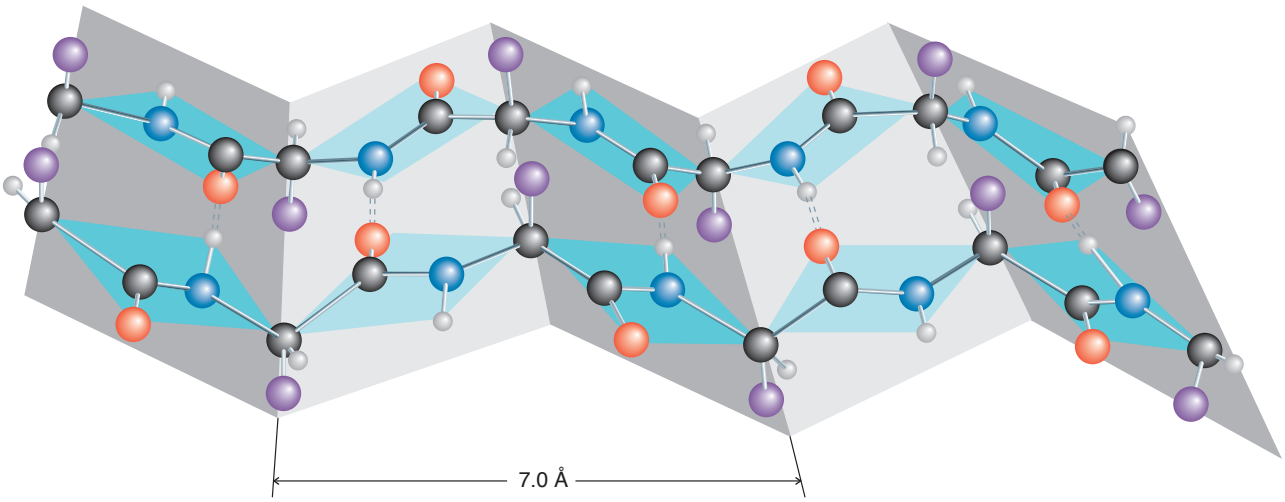


Figure 24.10 The β-pleated sheet or β configuration of a protein. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

- The ***α* helix** is also a very important secondary structure in proteins (Fig. 24.11).

The α-helix of a polypeptide is right-handed with 3.6 amino acid residues per turn. Each amide group in the chain has a hydrogen bond to an amide group at a distance of three amino acid residues in either direction, and the R groups all extend away from the axis of the helix. The repeat distance of the α helix is 5.4 Å.

The α-helical structure is found in many proteins; it is the predominant structure of the polypeptide chains of fibrous proteins such as *myosin,* the protein of muscle, and of *α-keratin,* the protein of hair, unstretched wool, and nails.

Helices and pleated sheets account for only about one-half of the structure of the average globular protein. The remaining polypeptide segments have what is called a **coil** or **loop conformation.** These nonrepetitive structures are not random; they are just more difficult to describe. Globular proteins also have stretches, called **reverse turns** or **β bends,** where the polypeptide chain abruptly changes direction. These often connect successive strands of β sheets and almost always occur at the surface of proteins.

Figure 24.12 shows the structure of the enzyme human carbonic anhydrase, based on X-ray crystallographic data. Segments of α helix (magenta) and β sheets (yellow) intervene between reverse turns and nonrepetitive structures (blue and white, respectively).

- The locations of the side chains of amino acids of globular proteins are usually those that we would expect from their polarities:

1. Residues with **nonpolar, hydrophobic side chains,** such as *valine, leucine, isoleucine, methionine, and phenylalanine,* are almost always found in the interior of the protein, out of contact with the aqueous solvent. (These hydrophobic interactions are largely responsible for the tertiary structure of proteins that we discuss in Section 24.8B.)
2. Side chains of **polar residues with positive or negative charges,** such as *arginine, lysine, aspartic acid, and glutamic acid,* are usually on the surface of the protein in contact with the aqueous solvent.
3. **Uncharged polar side chains,** such as those of *serine, threonine, asparagine, glutamine, tyrosine, and tryptophan,* are most often found on the surface, but some of these are found in the interior as well. When they are found in the interior, they are virtually all hydrogen bonded to other similar residues. Hydrogen bonding apparently helps neutralize the polarity of these groups.

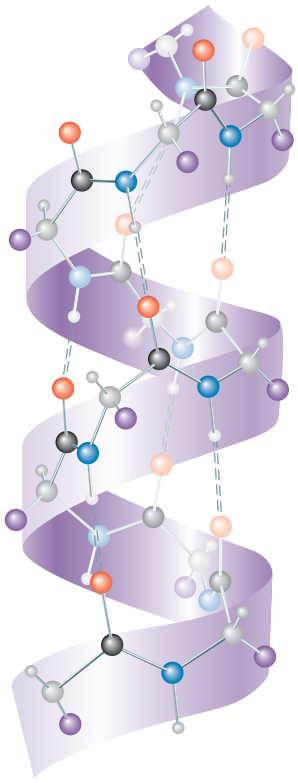


Figure 24.11 A representation of the α-helical structure of a polypeptide. Hydrogen bonds are denoted by dashed lines. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

Certain peptide chains assume what is called a **random coil arrangement**, a structure that is flexible, changing, and statistically random. Synthetic polylysine, for example, exists as a random coil and does not normally form an α helix. At pH 7, the ε-amino groups of the lysine residues are positively charged, and, as a result, repulsive forces between them are so large that they overcome any stabilization that would be gained through hydrogen bond formation of an α helix. At pH 12, however, the ε-amino groups are uncharged and polylysine spontaneously forms an α helix.

The presence of proline or hydroxyproline residues in polypeptide chains produces another striking effect: Because the nitrogen atoms of these amino acids are part of five-

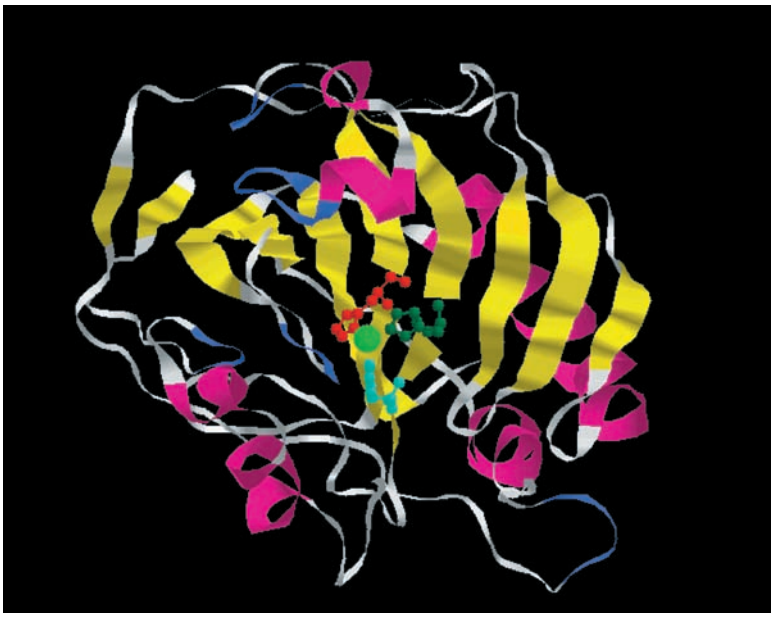


Figure 24.12 The structure of the enzyme human carbonic anhydrase, based on X-ray crystallographic data. Alpha helices are shown in magenta and strands of β-pleated sheets are yellow. Turns are shown in blue and random coils are white. The side chains of three histidine residues (shown in red, green, and cyan) coordinate with a zinc atom (light green). Not obvious from this image is the interesting fact that the C-terminus is tucked through a loop of the polypeptide chain, making carbonic anhydrase a rare example of a native protein in which the polypeptide chain forms a knot. (PDB ID CA2, http://www.pdb.org. Eriksson, Jones, Liljas, *Proteins: Structure, Function and Genetics,* **Volume 4**, Issue 4, 1988, pp. 274–282.)

membered rings, the groups attached by the nitrogen–α carbon bond cannot rotate enough to allow an α-helical structure. Wherever proline or hydroxyproline occur in a peptide chain, their presence causes a kink or bend and interrupts the α helix.

24.8B Tertiary Structure

- The **tertiary structure** of a protein is the overall three-dimensional shape that arises from all of the secondary structures of its polypeptide chain.

Proteins typically have either **globular** or **fibrous** tertiary structures. These tertiary structures do not occur randomly. Under the proper environmental conditions the tertiary structure of a protein occurs in one particular way—a way that is characteristic of that particular protein and one that is often highly important to its function.

Various forces are involved in stabilizing tertiary structures, including the disulfide bonds of the primary structure.

- One characteristic of most proteins is that the folding takes place in such a way as to expose the maximum number of polar (hydrophilic) groups to the aqueous environment and enclose a maximum number of nonpolar (hydrophobic) groups within its interior.

The soluble globular proteins tend to be much more highly folded than fibrous proteins. Myoglobin (Fig. 24.13) is an example of a globular protein. However, fibrous proteins also have a tertiary structure; the α-helical strands of α-keratin, for example, are wound together into a "superhelix." The superhelix makes one complete turn for each 35 turns of the α helix. The tertiary structure does not end here, however. Even the superhelices can be wound together to give a ropelike structure of seven strands.

24.8C Quaternary Structure

Many proteins exist as stable and ordered noncovalent aggregates of more than one polypeptide chain. The overall structure of a protein having multiple subunits is called its **quaternary structure.** The quaternary structure of hemoglobin, for example, involves four subunits (see Section 24.12).

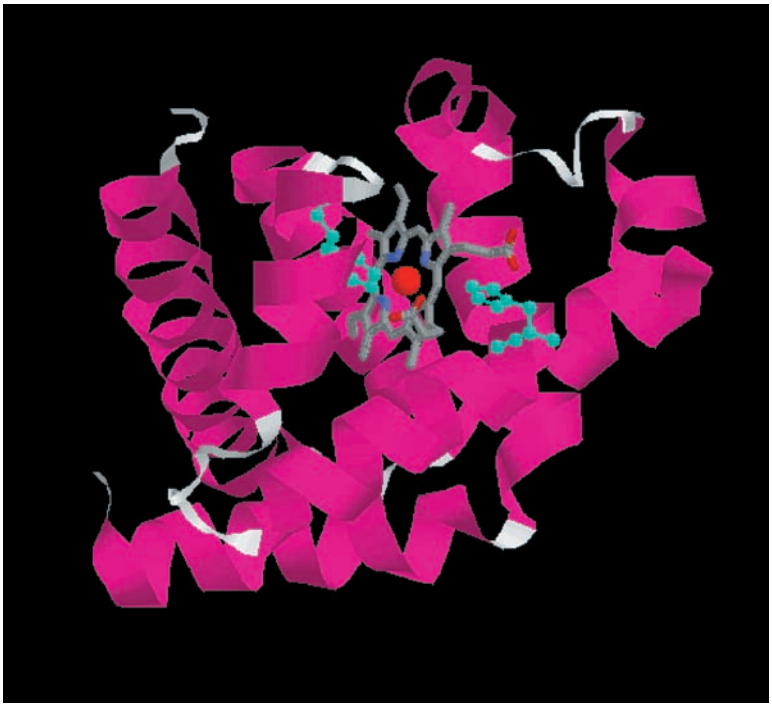


Figure 24.13 The three-dimensional structure of myoglobin. The heme ring is shown in gray. The iron atom is shown as a red sphere, and the histidine side chains that coordinate with the iron are shown in cyan. (PDB ID 1MBD, http://www.pdb.org. Phillips, S.E., Schoenberg, B.P. Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin. *Nature* 292, pp. 81–82, 1981.)

24.9 Introduction to Enzymes

- The reactions of cellular metabolism are mediated by remarkable biological catalysts called **enzymes.**

Enzymes have the ability to bring about vast increases in the rates of reactions; in most instances, the rates of enzyme-catalyzed reactions are faster than those of uncatalyzed reactions by factors of 10^6–10^{12}. For living organisms, rate enhancements of this magnitude are important because they permit reactions to take place at reasonable rates, even under the mild conditions that exist in living cells (i.e., approximately neutral pH and a temperature of about 35°C.)

- Enzymes show remarkable **specificity** for their **substrates** and for formation of specific products.

The specificity of enzymes is far greater than that shown by most chemical catalysts. In the enzymatic synthesis of proteins, for example (through reactions that take place on ribosomes, Section 25.5E), polypeptides consisting of well over 1000 amino acid residues are synthesized virtually without error. It was Emil Fischer's discovery, in 1894, of the ability of enzymes to distinguish between α- and β-glycosidic linkages (Section 22.12) that led him to formulate his **lock-and-key hypothesis** for enzyme specificity.

- According to the **lock-and-key hypothesis**, the specificity of an enzyme (the lock) and its substrate (the key) comes from their geometrically complementary shapes.
- In an enzyme-catalyzed reaction, the enzyme and the substrate combine to form an **enzyme–substrate complex.**
- Formation of the enzyme–substrate complex often induces a conformational change in the enzyme called an **induced fit** that allows it to bind the substrate more effectively.

Binding of the substrate can cause certain of its bonds to become strained, and therefore more easily broken. The product of the reaction usually has a different shape from the substrate, and this altered shape, or in some instances the intervention of another molecule, causes the complex to dissociate. The enzyme can then accept another molecule of the substrate, and the whole process is repeated:

$$\text{Enzyme} + \text{substrate} \rightleftharpoons \text{enzyme–substrate complex} \rightleftharpoons \text{enzyme} + \text{product}$$

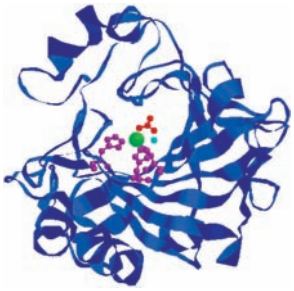
- The place where a substrate binds to an enzyme and where the reaction takes place is called the **active site.**

The noncovalent forces that bind the substrate to the active site are the same forces that account for the conformations of proteins: dispersion forces, electrostatic forces, hydrogen bonding, and hydrophobic interactions. The amino acids located in the active site are arranged so that they can interact specifically with the substrate.

- Reactions catalyzed by enzymes are **stereospecific** because enzymes are chiral.

The specificity of enzymes arises in the way enzymes bind their substrates. An α-glycosidase will only bind the α stereoisomeric form of a glycoside, not the β form. Enzymes that metabolize sugars bind only D sugars; enzymes that synthesize most proteins bind only L amino acids; and so on.

Although enzymes catalyze reactions stereospecifically, they often vary considerably in what is called their **geometric specificity.** By geometric specificity, we mean a specificity that is related to the identities of the chemical groups of the substrates. Some enzymes will accept only one compound as their substrate. Others, however, will accept a range of compounds with similar groups. Carboxypeptidase A, for example, will hydrolyze the C-terminal peptide from all polypeptides as long as the penultimate residue is not arginine, lysine, or proline and as long as the next preceding residue is not proline.

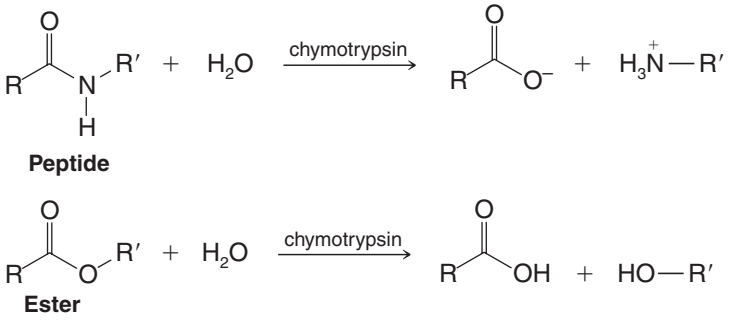


Carbonic anhydrase

Carbonic anhydrase is an enzyme that catalyzes the following reaction: $H_2O + CO_2 \rightleftharpoons H_2CO_3$. (PDB ID CA2, http://www.pdb.org. Eriksson, Jones, Liljas, *Proteins: Structure, Function and Genetics,* Volume 4, Issue 4, 1988, pp. 274–282.)

Certain RNA molecules, called ribozymes, can also act as enzymes. The 1989 Nobel Prize in Chemistry went to Sidney Altman (Yale University) and to Thomas R. Cech (University of Colorado, Boulder) for the discovery of ribozymes.

Chymotrypsin, a digestive enzyme that catalyzes the hydrolysis of peptide bonds, will also catalyze the hydrolysis of esters. We shall consider its mechanism of hydrolysis in Section 24.11.



Peptide **Ester**

- A compound that can negatively alter the activity of an enzyme is called an **inhibitor.**

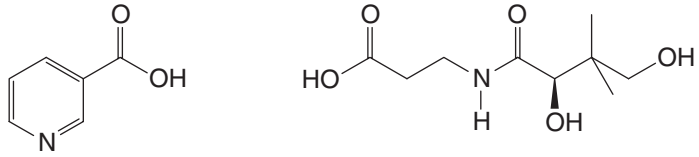
Helpful Hint

In *WileyPLUS* materials we have highlighted several coenzymes because they are the "organic chemistry machinery" of some enzymes. For example, see "The Chemistry of ... Pyridoxal Phosphate" and "The Chemistry of ... Thiamine."

A **competitive inhibitor** is a compound that competes directly with the substrate for the active site. We learned in Section 20.9, for example, that sulfanilamide is a competitive inhibitor for a bacterial enzyme that incorporates *p*-aminobenzoic acid into folic acid.

Some enzymes require the presence of a **cofactor.** The cofactor may be a metal ion as, for example, the zinc atom of human carbonic anhydrase (see the Chemistry of . . . box, Section 24.10 and Fig. 24.12). Others may require the presence of an organic molecule, such as NAD^+ (Section 14.10), called a **coenzyme.** Coenzymes become chemically changed in the course of the enzymatic reaction. NAD^+ becomes converted to **NADH.** In some enzymes the cofactor is permanently bound to the enzyme, in which case it is called a **prosthetic group.**

Many of the water-soluble vitamins are the precursors of coenzymes. Niacin (nicotinic acid) is a precursor of NAD^+, for example. Pantothenic acid is a precursor of coenzyme A.



Niacin **Pantothenic acid**

24.10 *Lysozyme: Mode of Action of an Enzyme*

Lysozyme is an enzyme that breaches the cell wall of gram-positive bacteria by hydrolyzing specific acetal linkages in the cell's peptidoglycan polymer, causing lysis and cell death. We shall discuss the mechanism of this reaction below, but first let us consider the structure of lysozyme. The primary structure of lysozyme is shown in Figure 24.14.

Lysozyme's secondary structure includes α-helices at residues 5–15, 24–34, and 88–96; β-pleated sheets involving residues 41–45 and 50–54; and a hairpin turn at residues 46–49. The remaining polypeptide segments of lysozyme have coil or loop formations. Glu-35 and Asp-52 are the amino acid residues directly involved in the hydrolysis reaction catalyzed by lysozyme. A three-dimensional structure of lysozyme is shown in Fig. 24.15. The amino acid residues responsible for its catalytic activity are highlighted in ball-and-stick format (Glu-35 and Asp-52 to the left).

We shall discuss the reaction catalyzed by lysozyme shortly; first however, the discovery of lysozyme is an interesting story in itself.

> One day in 1922 Alexander Fleming was suffering from a cold. This is not unusual in London, but Fleming was a most unusual man and he took advantage of the cold in a characteristic way. He allowed a few drops of his nasal mucus to fall on a

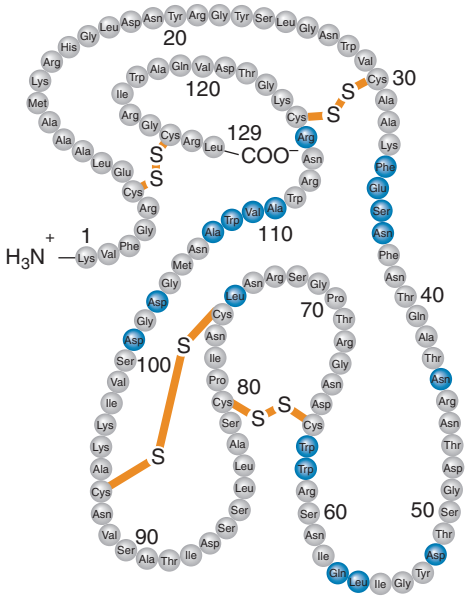


Figure 24.14 The primary structure of hen egg white lysozyme. The amino acids that line the substrate-binding pocket are shown in blue. (Reprinted with permission of John Wiley & Sons, Inc. from Voet, D. and Voet, J. G. *Biochemistry*, Second Edition. © 1995 Voet, D. and Voet, J. G.)

culture of bacteria he was working with and then put the plate to one side to see what would happen. Imagine his excitement when he discovered some time later that the bacteria near the mucus had dissolved away. For a while he thought his ambition of finding a universal antibiotic had been realized. In a burst of activity he quickly established that the antibacterial action of the mucus was due to the presence of an enzyme; he called this substance lysozyme because of its capacity to lyse, or dissolve, the bacterial cells. Lysozyme was soon discovered in many tissues and secretions of the human body, in plants, and most plentifully of all in the white of an egg. Unfortunately Fleming found that it is not effective against the most harmful bacteria. He had to wait 7 years before a strangely similar experiment revealed the existence of a genuinely effective antibiotic: penicillin.

This story was related by Professor David C. Phillips of Oxford University, who first elucidated the three-dimensional structure of lysozyme using X-ray crystallography.*

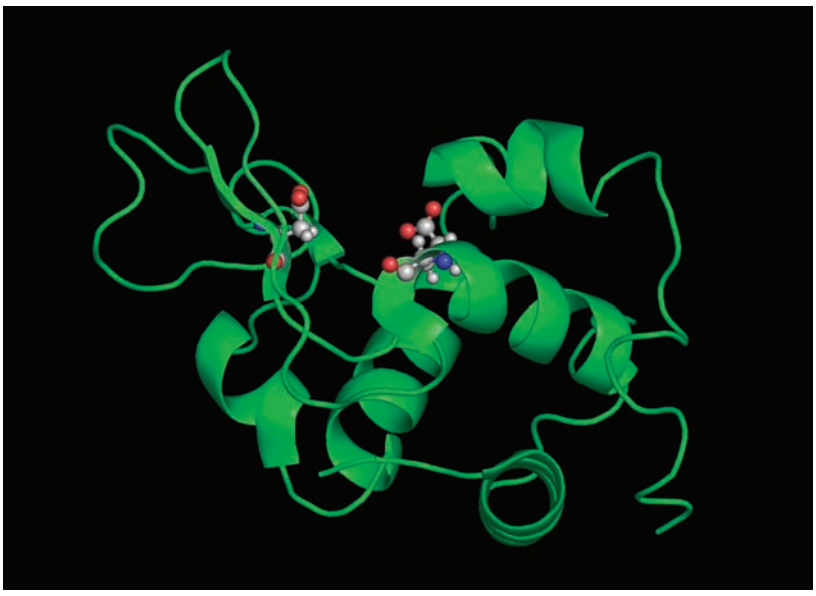


Figure 24.15 A ribbon diagram of lysozyme highlighting aspartic acid 52 (left) and glutamic acid 35 (right) in ball-and-stick format. (PDB ID: 1AZF, http://www.pdb.org. Lim, K., Nadarajah A., Forsythe, E.L., Pusey, M.L. Locations of bromide ions in tetragonal lysozyme crystals. *Acta Crystallogr.*, Sect. D, **54,** pp. 899–904, 1998.)

*Quotation from "The Three-Dimensional Structure of an Enzyme Molecule" by David C. Phillips. © 1966 by Scientific American, Inc. All rights reserved.

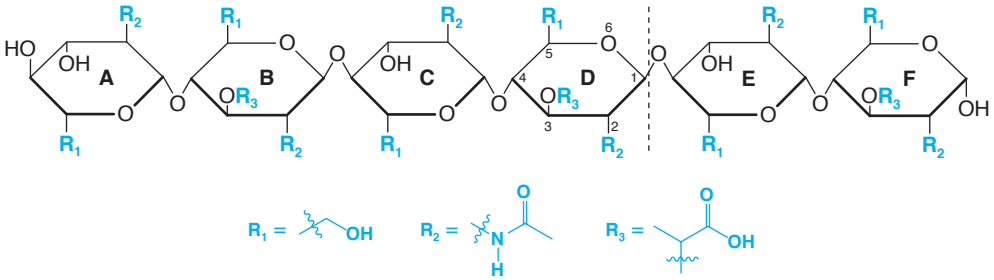
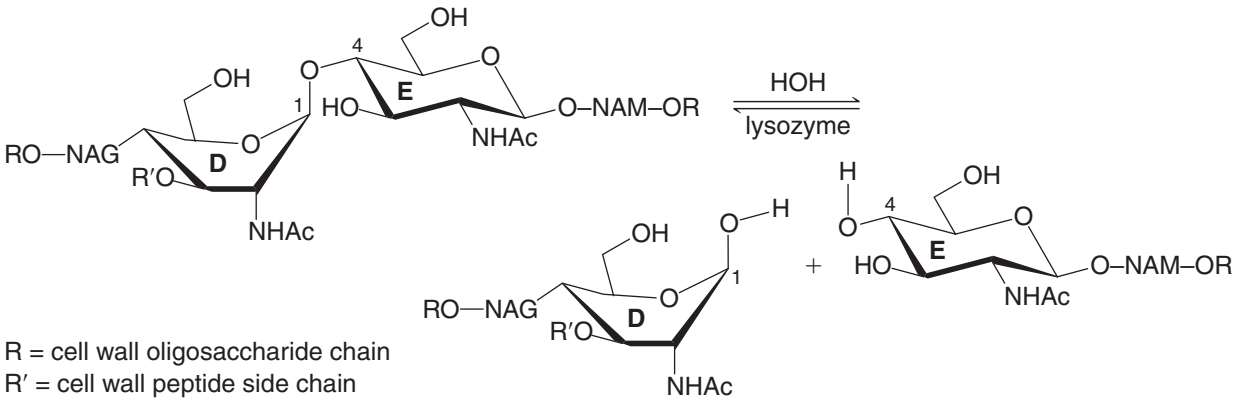


Figure 24.16 A hexasaccharide that has the same general structure as the cell wall polysaccharide on which lysozyme acts. Two different amino sugars are present: rings A, C, and E are derived from a monosaccharide called *N*-acetylglucosamine; rings B, D, and F are derived from a monosaccharide called *N*-acetylmuramic acid. When lysozyme acts on this oligosaccharide, hydrolysis takes place and results in cleavage at the glycosidic linkage between rings D and E.

As mentioned, lysozyme hydrolyzes glycosidic linkages in the peptidoglycan polymer of gram-positive bacterial cell walls. The structure of an oligosaccharide similar to the polysaccharide found in bacterial cell walls is shown in Fig. 24.16. *N*-Acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) form alternating repeat units in this polysaccharide.

Lysozyme selectively binds a six-unit segment of the peptidoglycan polymer and hydrolyzes specifically the acetal linkage between rings D and E shown in Fig. 24.16 (NAM and NAG units, respectively).

The overall reaction catalyzed by lysozyme is as follows:



Lysozyme binds the cell wall substrate in a cleft within its tertiary structure, such that the Glu-35 residue is close to the substrate on one side and Asp-52 is close on the other. Both amino acid residues are positioned in a way that facilitates reaction with the D–E glycosidic linkage of the polysaccharide.

Strong evidence from mass spectrometry suggests that the mechanism of lysozyme involves sequential S_N2 reactions and a covalent enzyme–substrate intermediate (based on work by Stephen Withers and colleagues at the University of British Columbia and elsewhere). Asp-52 acts as the nucleophile in the first step, covalently bonding the substrate to the enzyme. A water molecule acts as a nucleophile in the second step to complete the formation of product and free the substrate from the active site. In both steps, Glu-35 serves as a general acid–base catalyst. The details are as follows.

As lysozyme binds the substrate, the active site cleft closes slightly and C1 of ring D in the oligosaccharide substrate moves downward. The carboxylate group of Asp-52 attacks C1 of ring D from below (Figure 24.17), displacing the ring E C4 oxygen as a leaving group. The ring E C4 oxygen departs as a neutral species because it is protonated concurrently by the carboxylic acid group of Glu-35. The transition state for this S_N2 reaction is presumed to be the point at which ring D is nearly flat during the boat to chair conformational change.

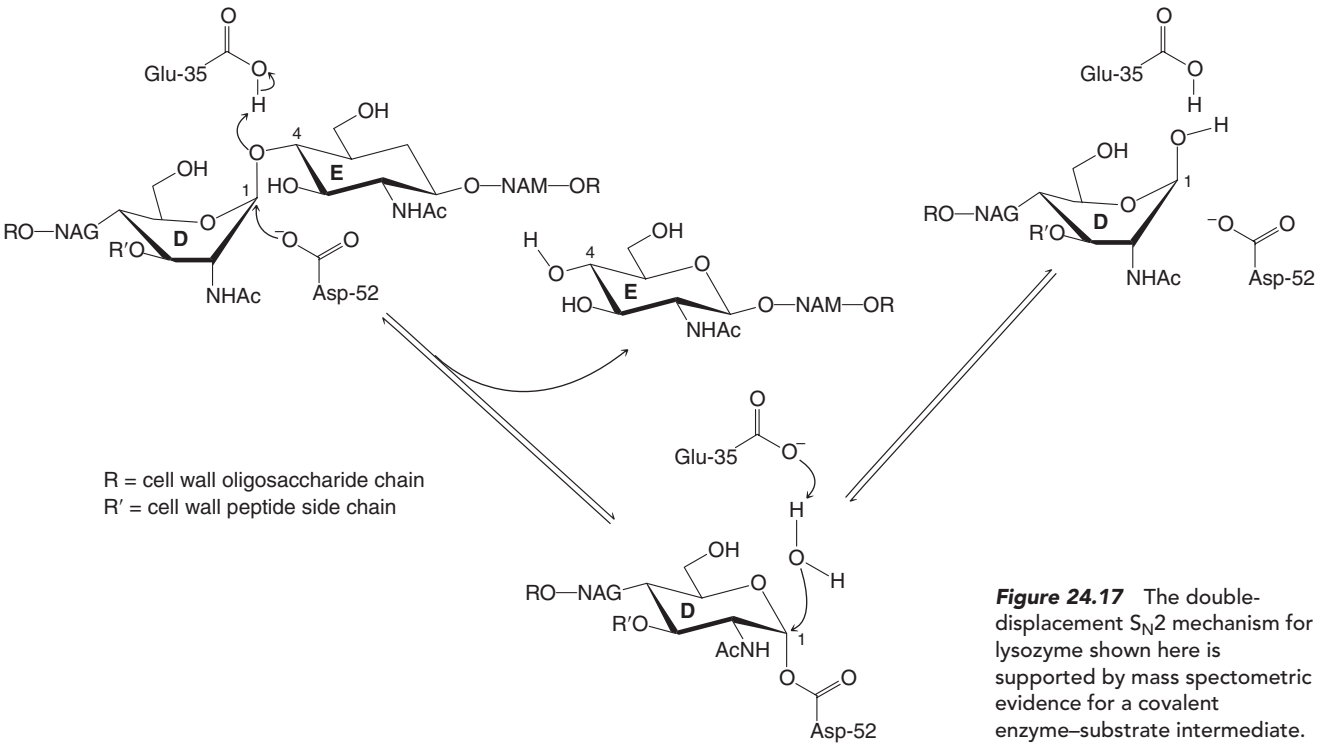
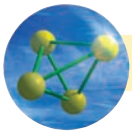


Figure 24.17 The double-displacement S_N2 mechanism for lysozyme shown here is supported by mass spectometric evidence for a covalent enzyme–substrate intermediate.

This step occurs with inversion, as expected for an S_N2 reaction, and leaves one part of the substrate covalently bound to the enzyme.

In the second step, a water molecule, now in the site formerly occupied by ring E, attacks C1 and displaces the carboxylate group of Asp-52 as a leaving group. The Glu-35 anion assists as a base by removing a proton from the water molecule as it bonds with C1 of ring D. The entire lysozyme molecule serves as the leaving group. This event also occurs with inversion, liberates the substrate from the active site, and returns lysozyme to readiness for another catalytic cycle. The overall mechanism is shown in Fig. 24.17.



THE CHEMISTRY OF . . .

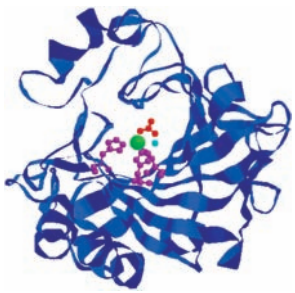
Carbonic Anhydrase: Shuttling the Protons

An enzyme called carbonic anhydrase regulates the acidity (pH) of blood and the physiological conditions relating to blood pH. The reaction that carbonic anhydrase catalyzes is the equilibrium conversion of water and carbon dioxide to carbonic acid (H_2CO_3).

$$H_2O + CO_2 \overset{\text{carbonic anhydrase}}{\rightleftharpoons} H_2CO_3 \rightleftharpoons HCO_3^- + H^+$$

The rate at which one breathes, for example, is influenced by one's relative blood acidity. Mountain climbers going to high elevations sometimes take a drug called Diamox (acetazolamide) to prevent altitude sickness. Diamox inhibits carbonic anhydrase, and this, in turn, increases blood acidity. This increased blood acidity stimulates breathing and thereby decreases the likelihood of altitude sickness.

Carbonic anhydrase consists of a chain of 260 amino acids that naturally folds into a specific globular shape. Included in its structure is a cleft or pocket, the active site, where the reactants are converted to products. The protein chain of carbonic anhydrase is shown here as a blue ribbon.

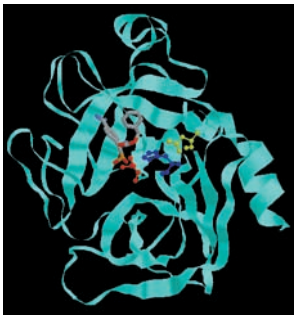


Carbonic anhydrase

At the active site of carbonic anhydrase a water molecule loses a proton to form a hydroxide (OH^-) ion. This proton is removed by a part of carbonic anhydrase that acts as a base. Ordinarily the proton of a water molecule is not very acidic. However, the Lewis acid–base interaction between a zinc cation at the active site of carbonic anhydrase and the oxygen atom of a water molecule leads to

positive charge on the water oxygen atom. This makes the protons of the water molecule more acidic. Removal of one of the protons of the water molecule forms hydroxide, which reacts with a carbon dioxide molecule at the active site to form HCO_3^- (hydrogen carbonate, or bicarbonate). In the structure of carbonic anhydrase shown here (based on X-ray crystallographic data), a bicarbonate ion at the active site is shown in red, the zinc cation at the active site is green, a water molecule is shown in blue, and the basic sites that coordinate with the zinc cation (as Lewis bases) or remove the proton from water to form hydroxide (as Brønsted–Lowry bases) are magenta (these bases are nitrogen atoms from histidine imidazole rings). No hydrogen atoms are shown in any of these species. As you can see, a remarkable orchestration of Lewis and Brønsted–Lowry acid–base reactions is involved in catalysis by carbonic anhydrase.

24.11 *Serine Proteases*



A serine protease

Chymotrypsin, trypsin, and elastin are digestive enzymes secreted by the pancreas into the small intestine to catalyze the hydrolysis of peptide bonds. These enzymes are all called **serine proteases** because the mechanism for their proteolytic activity (one that they have in common) involves a particular serine residue that is essential for their enzymatic activity. As another example of how enzymes work, we shall examine the mechanism of action of chymotrypsin.

Chymotrypsin is formed from a precursor molecule called chymotrypsinogen, which has 245 amino acid residues. Cleavage of two dipeptide units of chymotrypsinogen produces chymotrypsin. Chymotrypsin folds in a way that brings together histidine at position 57, aspartic acid at position 102, and serine at position 195. Together, these residues constitute what is called the **catalytic triad** of the active site (Fig. 24.18). Near the active site is a hydrophobic binding site, a slotlike pocket that preferentially accommodates the nonpolar side chains of Phe, Tyr, and Trp.

After chymotrypsin has bound its protein substrate, the serine residue at position 195 is ideally situated to attack the acyl carbon of the peptide bond (Fig. 24.19). This serine residue is made more nucleophilic by transferring its proton to the imidazole nitrogen of the histidine residue at position 57. The imidazolium ion that is formed is stabilized by the polarizing effect of the carboxylate ion of the aspartic acid residue at position 102. (Neutron diffraction studies, which show the positions of hydrogen atoms, confirm that the carboxylate ion remains as a carboxylate ion throughout and does not actually accept a proton

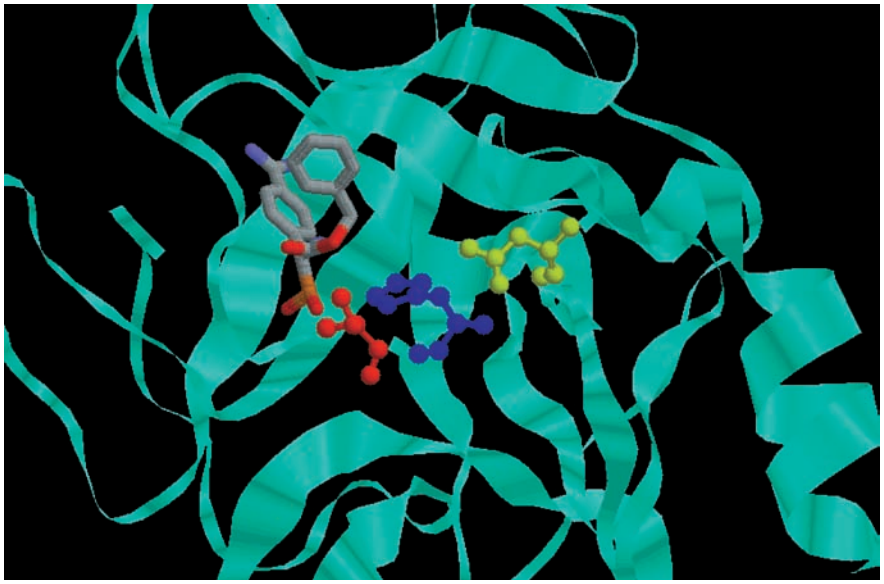


Figure 24.18 The catalytic triad in this serine protease (trypsin) is highlighted using the ball-and-stick model format for aspartic acid 52 (yellow-green), histidine 102 (purple), and serine 195 (red). A phosphonate inhibitor bound at the active site is shown in tube format. (This image and that in the margin, PDB ID: 1MAX, http://www.pdb.org, Bertrand, J.A., Oleksyszyn, J., Kam, C.M., Boduszek, B., Presnell, S., Plaskon, R.R., Suddath, F.L., Powers, J.C., Williams, L.D., Inhibition of trypsin and thrombin by amino(4-amidinophenyl)methanephosphonate diphenyl ester derivatives: X-ray structures and molecular models. *Biochemistry* **35**, pp. 3147–3155, 1996.)

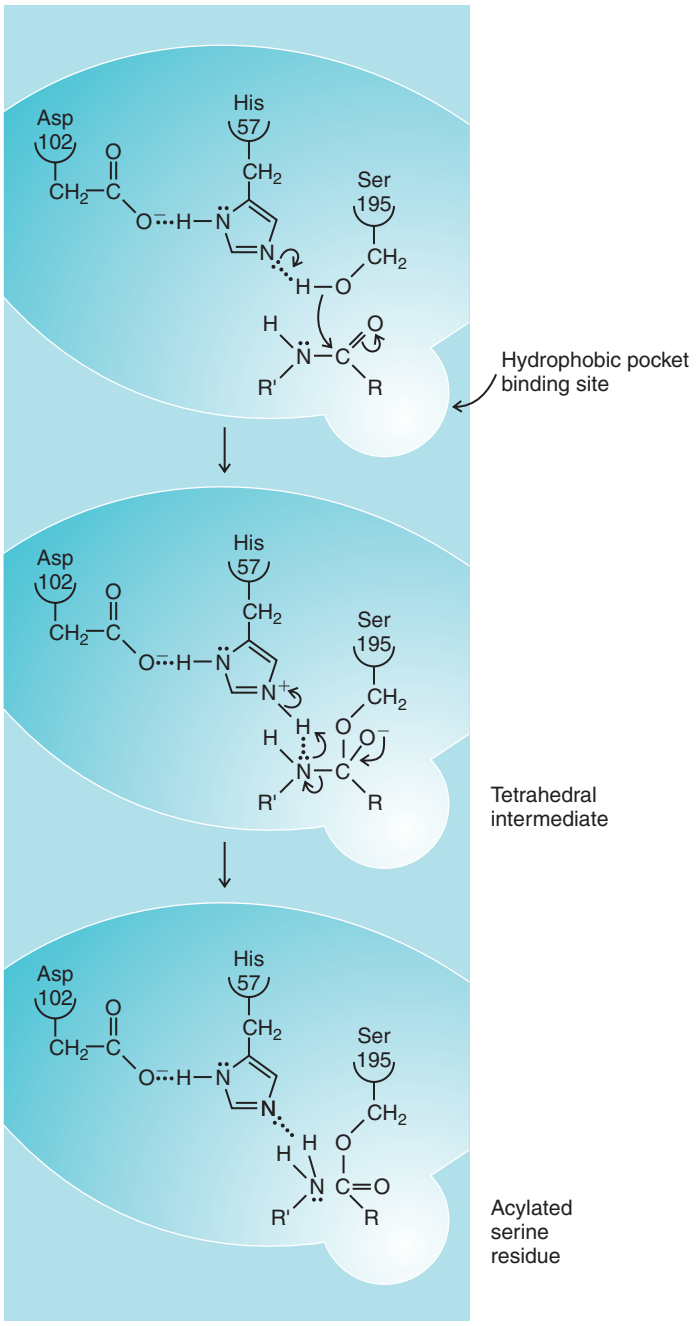


Figure 24.19 The catalytic triad of chymotrypsin causes cleavage of a peptide bond by acylation of serine residue 195 of chymotrypsin. Near the active site is a hydrophobic binding site that accommodates nonpolar side chains of the protein.

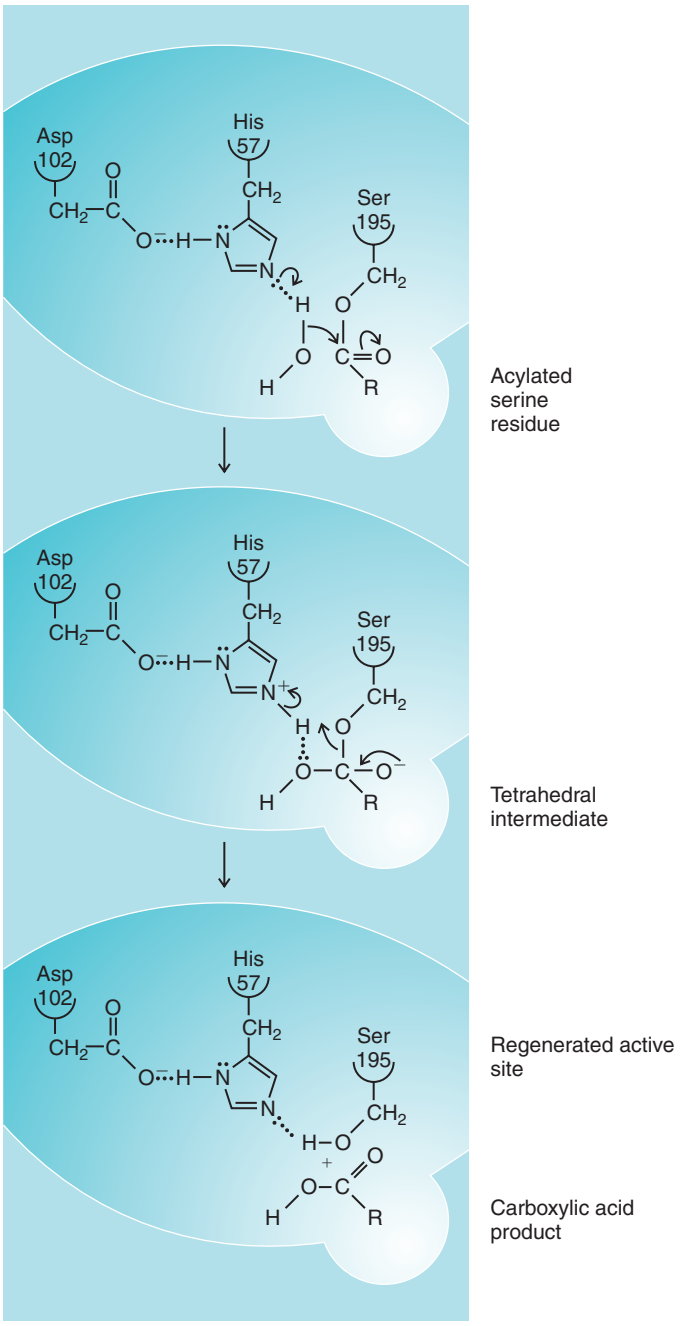


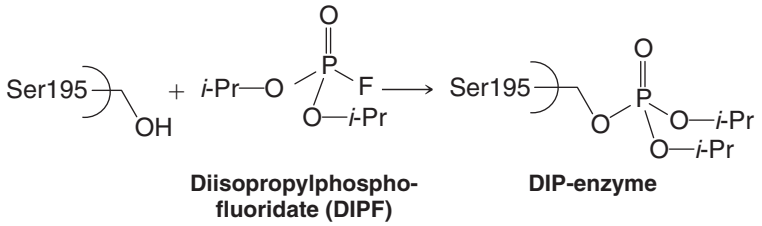
Figure 24.20 Regeneration of the active site of chymotrypsin. Water causes hydrolysis of the acyl–serine bond.

from the imidazole.) Nucleophilic attack by the serine leads to an acylated serine through a tetrahedral intermediate. The new N-terminal end of the cleaved polypeptide chain diffuses away and is replaced by a water molecule.

Regeneration of the active site of chymotrypsin is shown in Fig. 24.20. In this process water acts as the nucleophile and, in a series of steps analogous to those in Fig. 24.19, hydrolyzes the acyl–serine bond. The enzyme is now ready to repeat the whole process.

There is much evidence for this mechanism that, for reasons of space, we shall have to ignore. One bit of evidence deserves mention, however. There are compounds such as

diisopropylphosphofluoridate (DIPF) that irreversibly inhibit serine proteases. It has been shown that they do this by reacting only with Ser 195:



Recognition of the inactivating effect of DIPF came about as a result of the discovery that DIPF and related compounds are powerful **nerve poisons.** (They are the "nerve gases" of military use, even though they are liquids dispersed as fine droplets, and not gases.) Diisopropylphosphofluoridate inactivates **acetylcholinesterase** (Section 20.3) by reacting with it in the same way that it does with chymotrypsin. Acetylcholinesterase is a **serine esterase** rather than a serine protease.

24.12 *Hemoglobin: A Conjugated Protein*

Some proteins, called **conjugated proteins,** contain as a part of their structure a nonprotein group called a **prosthetic group.** An example is the oxygen-carrying protein hemoglobin. Each of the four polypeptide chains of hemoglobin is bound to a prosthetic group called *heme* (Fig. 24.21). The four polypeptide chains of hemoglobin are wound in such a way as to give hemoglobin a roughly spherical shape (Fig. 24.22). Moreover, each heme group lies in a crevice with the hydrophobic vinyl groups of its porphyrin structure surrounded by hydrophobic side chains of amino acid residues. The two propanoate side chains of heme lie near positively charged amino groups of lysine and arginine residues.

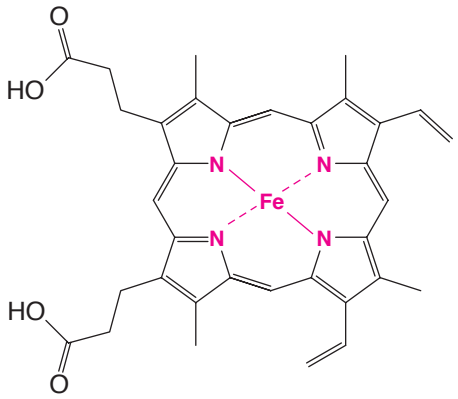


Figure 24.21 The structure of heme, the prosthetic group of hemoglobin. Heme has a structure similar to that of chlorophyll (Fig. 22.1) in that each is derived from the heterocyclic ring, porphyrin. The iron of heme is in the ferrous (2+) oxidation state.

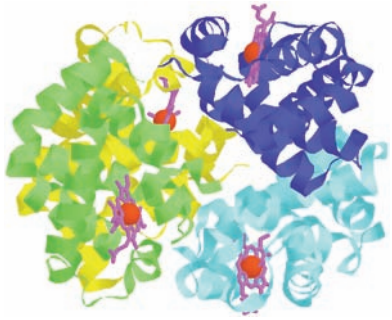


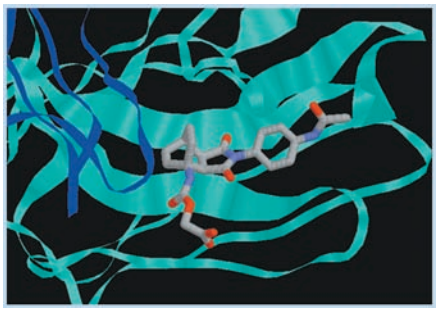
Figure 24.22 Hemoglobin. The two α subunits of hemoglobin are shown in blue and green. The two β subunits are shown in yellow and cyan. The four heme groups are shown in purple, and their iron atoms are in red. (PDB ID: IOUU, http://www.pdb.org. Tame, J.R., Wilson, J.C., Weber, R.E. The crystal structures of trout Hb I in the deoxy and carbonmonoxy forms. *J. Mol. Biol.* **259**, pp. 749–760, 1996.)

THE CHEMISTRY OF . . .

Some Catalytic Antibodies

Antibodies are chemical warriors of the immune system. Each antibody is a protein produced specifically in response to an invading chemical species (e.g., molecules on the surface of a virus or pollen grain). The purpose of antibodies is to bind with these foreign agents and cause their removal from the organism. The binding of each antibody with its target (the antigen) is usually highly specific.

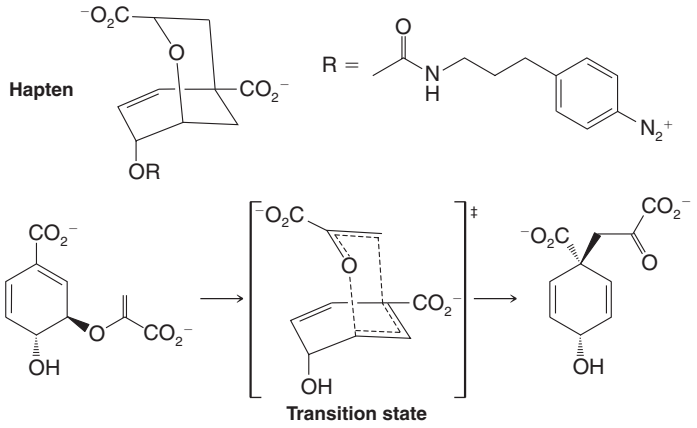
One way that *catalytic* antibodies have been produced is by prompting an immune response to a chemical species resembling the transition state for a reaction. According to this idea, if an antibody is created that preferentially binds with a stable molecule that has a transition state-like structure, other molecules that are capable of reaction *through* this transition state should, in principle, react faster as a result of binding with the antibody. (By facilitating association of the reactants and favoring formation of the transition state structure, the antibody acts in a way similar to an enzyme.) In stunning fashion, precisely this strategy has worked to generate catalytic antibodies for certain Diels–Alder reactions, Claisen rearrangements, and ester hydrolyses. Chemists have synthesized stable molecules that resemble transition states for these reactions, allowed antibodies to be generated against these molecules (called haptens), and then isolated the resulting antibodies. The antibodies thus produced are catalysts when actual substrate molecules are provided.



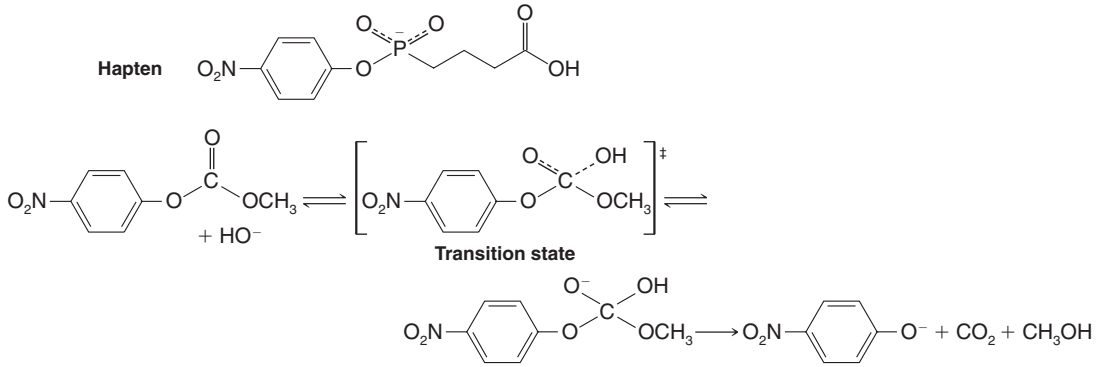
A hapten related to the Diels–Alder adduct from cyclohexadiene and maleimide, bound within a Diels–Alderase catalytic antibody. (PDB ID: 1A4K, http://www.pdb.org. Romesberg, F.E., Spiller, B., Schultz, P.G., Stevens, R.C. Immunological origins of binding and catalysis in a Diels–Alderase antibody. *Science* **279**, pp. 1929–1933, 1998.)

The following are examples of haptens used as transition state analogs to elicit catalytic antibodies for a Claisen rearrangement, hydrolysis of a carbonate, and a Diels–Alder reaction. The reaction catalyzed by the antibody generated from each hapten is shown as well.

Claisen Rearrangement

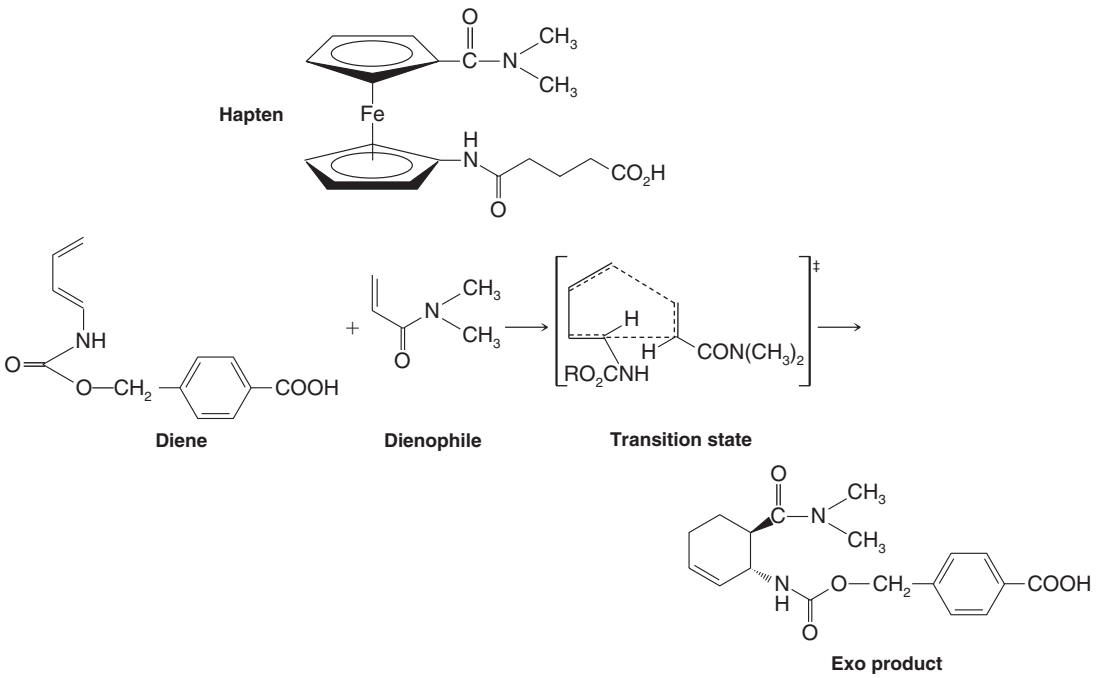


Carbonate Hydrolysis



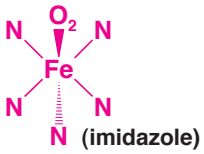
(continues on the next page)

Diels–Alder Reaction



This marriage of enzymology and immunology, resulting in chemical offspring, is just one area of exciting research at the interface of chemistry and biology.

The iron of the heme group is in the 2+ (ferrous) oxidation state and it forms a coordinate bond to a nitrogen of the imidazole group of histidine of the polypeptide chain. This leaves one valence of the ferrous ion free to combine with oxygen as follows:



A portion of oxygenated hemoglobin

The fact that the ferrous ion of the heme group combines with oxygen is not particularly remarkable; many similar compounds do the same thing. What is remarkable about hemoglobin is that when the heme combines with oxygen the ferrous ion does not become readily oxidized to the ferric state. Studies with model heme compounds in water, for example, show that they undergo a rapid combination with oxygen but they also undergo a rapid oxidation of the iron from Fe^{2+} to Fe^{3+}. When these same compounds are embedded in the hydrophobic environment of a polystyrene resin, however, the iron is easily oxygenated and deoxygenated, and this occurs *with no change in oxidation state of iron.* In this respect, it is especially interesting to note that X-ray studies of hemoglobin have revealed that the polypeptide chains provide each heme group with a similar hydrophobic environment.

24.13 Purification and Analysis of Polypeptides and Proteins

24.13A Purification

There are many methods used to purify polypeptides and proteins. The specific methods one chooses depend on the source of the protein (isolation from a natural source or chemical synthesis), its physical properties, including isoelectric point (p*I*), and the quantity of the protein on hand. Initial purification methods may involve precipitation, various forms of column chromatography, and electrophoresis. Perhaps the most important final method for peptide purification, HPLC, is used to purify both peptides generated by automated synthesis and peptides and proteins isolated from nature.

24.13B Analysis

A variety of parameters are used to characterize polypeptides and proteins. One of the most fundamental is molecular weight. Gel electrophoresis can be used to measure the approximate molecular weight of a protein. Gel electrophoresis involves migration of a peptide or protein dissolved in a buffer through a porous polymer gel under the influence of a high-voltage electric field. The buffer used (typically about pH 9) imparts an overall negative charge to the protein such that the protein migrates toward the positively charged terminal. Migration rate depends on the overall charge and size of the protein as well as the average pore size of the gel. The molecular weight of the protein is inferred by comparing the distance traveled through the gel by the protein of interest with the migration distance of proteins with known molecular weights used as internal standards. The version of this technique called SDS–PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) allows protein molecular weight determinations with an accuracy of about 5–10%.

Mass spectrometry can be used to determine a peptide's molecular weight with very high accuracy and precision. Earlier we discussed mass spectrometry in the context of protein sequencing. Now we shall consider the practical aspects of how molecules with very high molecular weight, such as proteins, can be transferred to the gas phase for mass spectrometric analysis. This is necessary, of course, whether the analysis regards peptide sequencing or full molecular analysis. Small organic molecules, as we discussed in Chapter 9, can be vaporized simply with high vacuum and heat. High-molecular-weight species cannot be transferred to the gas phase solely with heat and vacuum. Fortunately, very effective techniques have been developed for generating gas-phase ions of large molecules without destruction of the sample.

One ionization method is electrospray ionization (ESI, Fig. 24.23), whereby a solution of a peptide (or other analyte) in a volatile solvent containing a trace of acid is sprayed through a high-voltage nozzle into the vacuum chamber of a mass spectrometer. The acid in the solvent generates ions by protonating Lewis basic sites within the analyte. Peptides are typically protonated multiple times. Once injected through the high-voltage nozzle into the vacuum chamber, solvent molecules evaporate from the analyte ions (Fig. 24.23*a*), and the ions are drawn into the mass analyzer (Fig. 24.23*b*). The mass analyzer detects the analyte ions according to their time of flight, and registers their mass-to-charge ratio (*m/z*) (Fig. 24.23*c*). Each peak displayed in the mass spectrum represents the molecular weight of an ion divided by the number of positive charges it carries. From this series of *m/z* peaks, the molecular weight of the analyte is calculated by a computerized process called deconvolution. An example of a deconvoluted spectrum, indicating a molecular weight of 46,360 atomic mass units (daltons), is shown in Fig. 24.23*d*.

If fragmentation of the analyte molecules is desired, it can be caused by collision-induced dissociation (CID, Section 24.5E). In this case, tandem mass spectrometry is necessary because the first mass analyzer in the system is used to select fragments of the peptide from CID based on their overall mass, while the second mass analyzer in the system records the spectrum of the selected peptide fragment. Multiple fragments from the CID procedure can be analyzed this way. The final spectrum for each peptide fragment selected has the typical appearance of a family of ions, as shown below.

One-quarter of the 2002 Nobel Prize in Chemistry was awarded to John B. Fenn for his development of ESI mass spectrometry. Another quarter of the prize was awarded to Koichi Tanaku for discoveries that led to matrix-assisted laser desorption ionization (MALDI, see below).

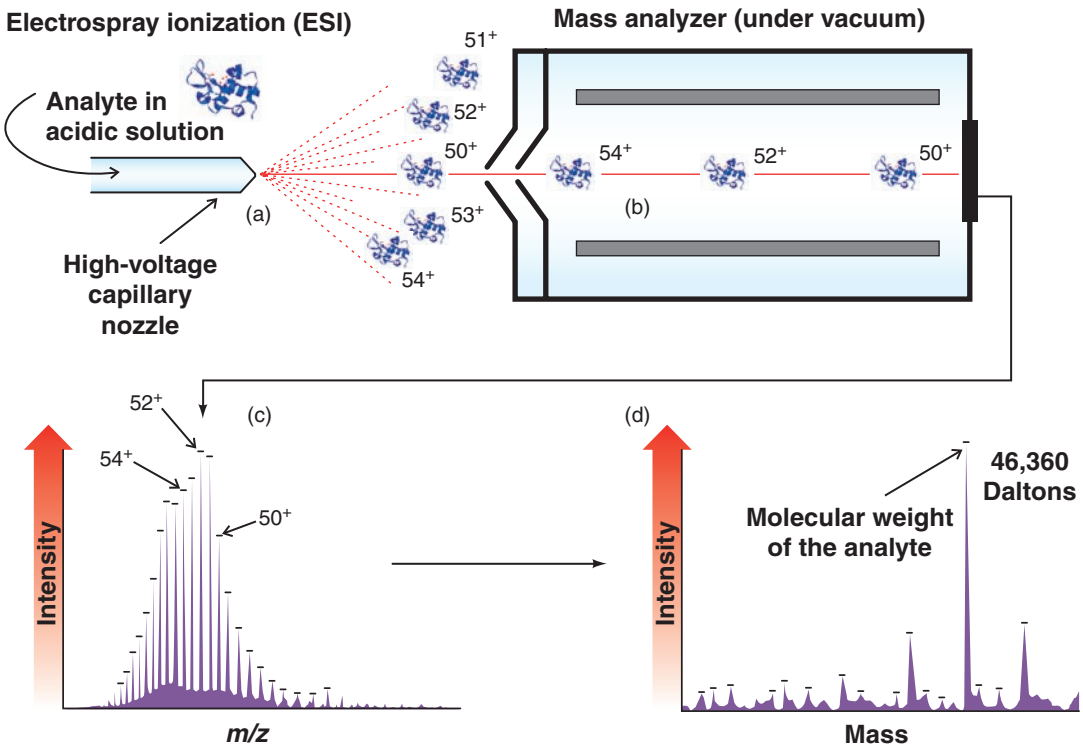


Figure 24.23 Electrospray ionization (ESI) mass spectrometry. (*a*) Analyte ions, protonated multiple times by an acidic solvent system, are sprayed through a high-voltage nozzle into a vacuum chamber (diagram is not to scale). Molecules of the solvent evaporate. The mutiply charged analyte ions are drawn into the mass analyzer. (*b*) The analyte ions are separated and detected in the mass analyzer. (*c*) The family of detected ions is displayed in a spectrum according to *m/z* ratio. (*d*) Computerized deconvolution of the *m/z* peak series leads to the molecular weight of the analyte.

Mass spectrometry with electrospray ionization (ESI-MS) is especially powerful when combined with HPLC because the two techniques can be used in tandem. With such an instrument the effluent from the HPLC is introduced directly into an ESI mass spectrometer. Thus, chromatographic separation of peptides in a mixture and direct structural information about each of them are possible using this technique.

Another method for ionization of nonvolatile molecules is MALDI (matrix-assisted laser desorption ionization, Section 9.18A). Energy from laser bombardment of a sample adsorbed in a solid chemical matrix leads to generation of gas-phase ions that are detected by the mass spectrometer. Both MALDI and ESI are common ionization techniques for the analysis of biopolymers.

24.14 Proteomics

Proteomics and genomics are two fields that have blossomed in recent years. **Proteomics** has to do with the study of all proteins that are expressed in a cell at a given time. **Genomics** (Sections 25.1 and 25.9) focuses on the study of the complete set of genetic instructions in an organism. While the genome holds the instructions for making proteins, it is proteins that carry out the vast majority of functions in living systems. Yet, compared to the tens of thousands of proteins encoded by the genome, we know the structure and function of only a relatively small percentage of proteins in the proteome. For this reason, the field of pro-

teomics has moved to a new level of importance since completion of sequencing the human genome. Many potential developments in health care and medicine now depend on identifying the myriad of proteins that are expressed at any given time in a cell, along with elucidation of their structures and biochemical function. New tools for medical diagnosis and targets for drug design will undoubtedly emerge at an increasing rate as the field of proteomics advances.

One of the basic challenges in proteomics is simply separation of all the proteins present in a cell extract. The next challenge is identification of those proteins that have been separated. Separation of proteins in cell extracts has classically been carried out using two-dimensional polyacrylamide gel electrophoresis (2D PAGE). In 2D PAGE the mixture of proteins extracted from an organism is separated in one dimension of the gel by the isoelectric point (a technique called isoelectric focusing) and in the second dimension by molecular weight. The result is a set of spots in the two-dimensional gel field that represents the location of separated proteins. The protein spots on the gel may then be extracted and analyzed by mass spectrometry or other methods, either as intact proteins or as enzymatic digests. Comparison of the results from mass spectrometry with protein mass spectrometry databases allows identification of many of the proteins separated by the gel.

There are limitations to protein separation by 2D PAGE, however. Not all proteins are amenable to 2D PAGE due to their size, charge, or specific properties. Furthermore, more than one protein may migrate to the same location if their isoelectric points and molecular weights are similar. Finally, 2D PAGE has inherent limits of detection that can leave some proteins of low concentration undetected.

An improvement over 2D PAGE involves two-dimensional microcapillary HPLC coupled with mass spectrometry (see Fig. 24.24). In this technique, called MudPIT (multidimensional protein identification technology, developed by John Yates and co-workers at Scripps Research Institute), a microcapillary HPLC column is used that has been packed first with a strong cation-exchange resin and then a reversed-phase (hydrophobic) material. The two packing materials used in sequence and with different resolving properties represent the two-dimensional aspect of this technique. A peptide mixture is introduced to the microcapillary column and eluted with pH and solvent gradients over a sequence of automated steps. As the separated peptides are eluted from the column they pass directly into a mass spectrometer. Mass spectrometric data obtained for each protein represent a signature that allows identification of the protein by comparison with a protein mass spectrometry database. This technique of 2D HPLC coupled with mass spectrometry is inherently more sensitive and general than 2D PAGE. One powerful example of its use is the identification by Yates and co-workers of nearly 1500 proteins from the *Saccharomyces cerevisiae* (baker's yeast) proteome in one integrated analysis.

Beyond the identification of proteins, quantitative measurement of the amounts of various proteins that are expressed is also important in proteomics. Various disease states or environmental conditions experienced by a cell may influence the amount of some proteins that are expressed. Quantitative tracking of these changes as a function of cell state could be relevant to studies of disease and the development of therapies. A technique using reagents called isotope-coded affinity tags (ICAT, developed at the University of Washington) allows quantitative analysis and identification of components in complex protein mixtures. The ICAT analysis involves mass spectrometric comparison of isotopically labeled and unlabeled protein segments that have been isolated by affinity chromatography and purified by microcapillary HPLC.

Hand in hand with identification and quantification of proteins remains the need to determine full three-dimensional protein structures. Even though thousands of proteins are encoded in the genome, only a relative handful of them have been studied in depth in terms of detailed structure and function. Full structure determination will therefore continue to be central to the field of proteomics. X-Ray crystallography, NMR, and mass spectrometry are key tools that will be applied ever more fervently as the quest intensifies to elucidate as many structures in the proteome as possible.

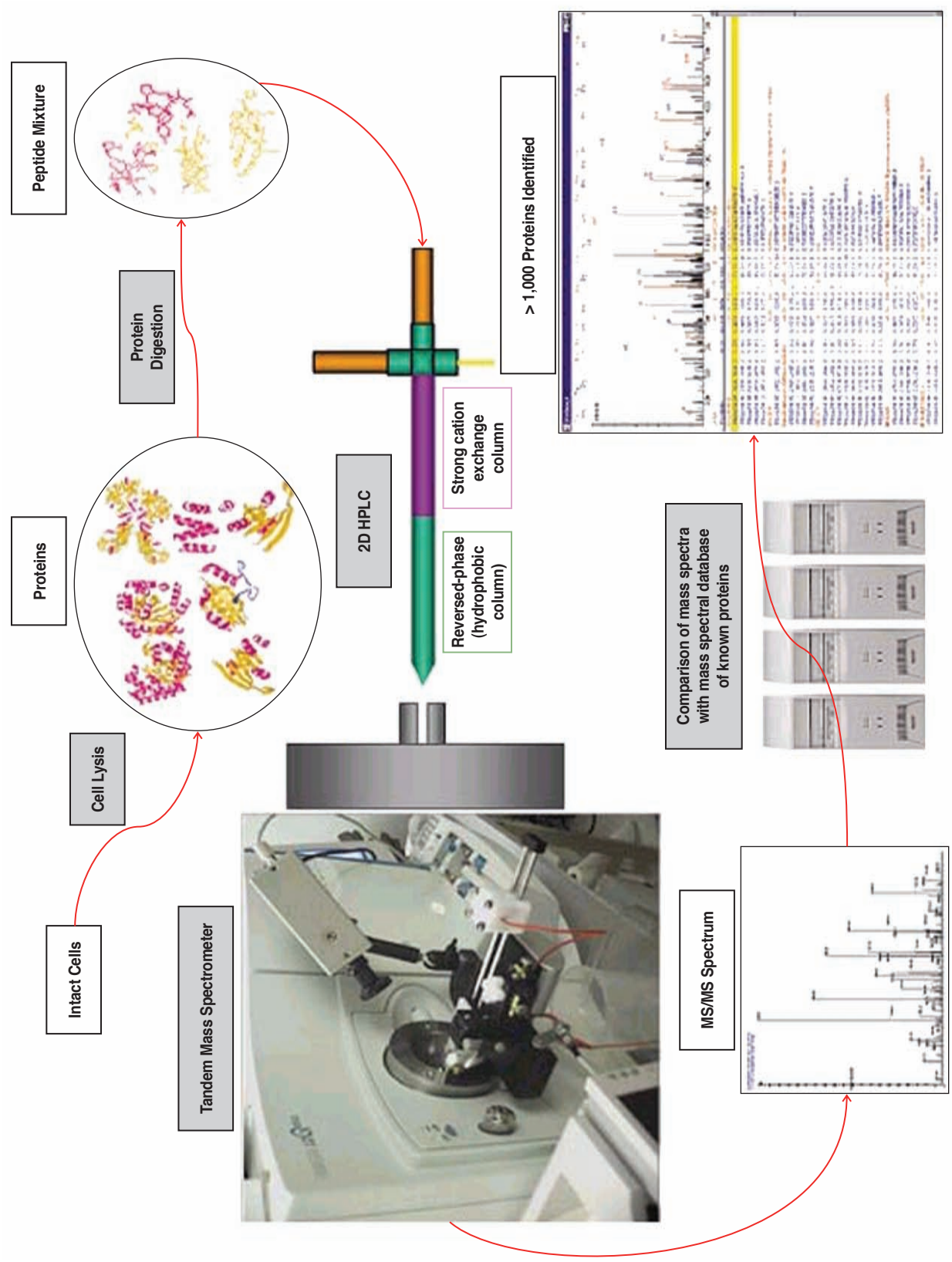


Figure 24.24 The high-throughput multidimensional protein identification technology (MudPIT) process. (Reprinted from *International Journal for Parasitology*, December 4, 32(13), Carucci, D.J.; Yates, J. R. 3rd; Florens, L.; Exploring the Proteome of Plasmodium, pp. 1539–1542, Copyright 2002, with permission from Elsevier.)

Key Terms and Concepts

The key terms and concepts that are highlighted in **bold, blue text** within the chapter are defined in the glossary (at the back of the book) and have hyperlinked definitions in the accompanying ***WileyPLUS*** course (www.wileyplus.com)

Problems

Note to Instructors: Many of the homework problems are available for assignment via ***WileyPLUS***, an online teaching and learning solution.

STRUCTURE AND REACTIVITY

24.17 **(a)** Which amino acids in Table 24.1 have more than one chirality center?

(b) Write Fischer projections for the isomers of each of these amino acids that would have the L configuration at the α carbon.

(c) What kind of isomers have you drawn in each case?

24.18 **(a)** What product would you expect to obtain from treating tyrosine with excess bromine water?

(b) What product would you expect to be formed in the reaction of phenylalanine with ethanol in the presence of hydrogen chloride?

(c) What product would you expect from the reaction of alanine and benzoyl chloride in aqueous base?

24.19 **(a)** On the basis of the following sequence of reactions, Emil Fischer was able to show that (−)-serine and L-(+)-alanine have the same configuration. Write Fischer projections for the intermediates **A–C**:

$$\text{(−)-Serine} \xrightarrow[CH_3OH]{HCl} \mathbf{A}\ (C_4H_{10}ClNO_3) \xrightarrow{PCl_5} \mathbf{B}\ (C_4H_9Cl_2NO_2) \xrightarrow[(2)\ OH^-]{(1)\ H_3O^+,\ H_2O,\ heat} \mathbf{C}\ (C_3H_6ClNO_2) \xrightarrow[\text{dilute } H_3O^+]{Na-Hg} \text{L-(+)-alanine}$$

(b) The configuration of L-(+)-cysteine can be related to that of L-(−)-serine through the following reactions. Write Fischer projections for **D** and **E**:

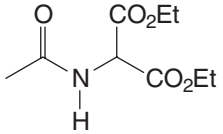
$$\mathbf{B}\ \text{(from part a)} \xrightarrow{OH^-} \mathbf{D}\ (C_4H_8ClNO_2) \xrightarrow{NaSH} \mathbf{E}\ (C_4H_9NO_2S) \xrightarrow[(2)\ OH^-]{(1)\ H_3O^+,\ H_2O,\ heat} \text{L-(+)-cysteine}$$

(c) The configuration of L-(−)-asparagine can be related to that of L-(−)-serine in the following way. What is the structure of **F**?

$$\text{L-(−)-Asparagine} \xrightarrow[\substack{\text{Hofmann}\\\text{rearrangement}}]{NaOBr/OH^-} \mathbf{F}\ (C_3H_7N_2O_2)$$

$$\mathbf{C}\ \text{(from part a)} \xrightarrow{NH_3} \mathbf{F}$$

24.20 **(a)** DL-Glutamic acid has been synthesized from diethyl acetamidomalonate in the following way. Outline the reactions involved.



Diethyl acetamidomalonate

$$+\ \text{CH}_2{=}\text{CHCN} \xrightarrow[\substack{EtOH\\(95\%\ yield)}]{NaOEt}$$

$$\mathbf{G}\ (C_{12}H_{18}N_2O_5) \xrightarrow[\substack{reflux\ 6\ h\\(66\%\ yield)}]{concd\ HCl} \text{DL-glutamic acid}$$

(b) Compound **G** has also been used to prepare the amino acid DL-ornithine through the following route. Outline the reactions involved here.

$$\mathbf{G}\ (C_{12}H_{18}N_2O_5) \xrightarrow[\substack{68°C,\ 1000\ psi\\(90\%\ yield)}]{H_2,\ Ni} \mathbf{H}\ (C_{10}H_{16}N_2O_4,\ \text{a } \delta\text{-lactam}) \xrightarrow[\substack{reflux\ 4\ h\\(97\%\ yield)}]{concd\ HCl}$$

DL-ornithine hydrochloride ($C_5H_{13}ClN_2O_2$)

(L-Ornithine is a naturally occurring amino acid but does not occur in proteins. In one metabolic pathway L-ornithine serves as a precursor for L-arginine.)

24.21 Synthetic polyglutamic acid exists as an α helix in solution at pH 2–3. When the pH of such a solution is gradually raised through the addition of a base, α dramatic change in optical rotation takes place at pH 5. This change has been associated with the unfolding of the α helix and the formation of a random coil. What structural feature of polyglutamic acid and what chemical change can you suggest as an explanation of this transformation?

PEPTIDE SEQUENCING

24.22 Bradykinin is a nonapeptide released by blood plasma globulins in response to a wasp sting. It is a very potent pain-causing agent. Its constituent amino acids are 2R, G, 2F, 3P, S. The use of 2,4-dinitrofluorobenzene and carboxypeptidase shows that both terminal residues are arginine. Partial acid hydrolysis of bradykinin gives the following di- and tripeptides:

FS + PGF + PP + SPF + FR + RP

What is the amino acid sequence of bradykinin?

24.23 Complete hydrolysis of a heptapeptide showed that it has the following constituent amino acids:

2A, E, L, K, F, V

Deduce the amino acid sequence of this heptapeptide from the following data.

1. Treatment of the heptapeptide with 2,4-dinitrofluorobenzene followed by incomplete hydrolysis gave, among other products: valine labeled at the α-amino group, lysine labeled at the ε-amino group, and a dipeptide, DNP—VL (DNP = 2,4-dinitrophenyl-).
2. Hydrolysis of the heptapeptide with carboxypeptidase gave an initial high concentration of alanine, followed by a rising concentration of glutamic acid.
3. Partial enzymatic hydrolysis of the heptapeptide gave a dipeptide **(A)** and a tripeptide **(B).**
   - **a.** Treatment of **A** with 2,4-dinitrofluorobenzene followed by hydrolysis gave DNP-labeled leucine and lysine labeled only at the ε-amino group.
   - **b.** Complete hydrolysis of **B** gave phenylalanine, glutamic acid, and alanine. When **B** was allowed to react with carboxypeptidase, the solution showed an initial high concentration of glutamic acid. Treatment of **B** with 2,4-dinitrofluorobenzene followed by hydrolysis gave labeled phenylalanine.

Challenge Problem

24.24 Part of the evidence for restricted rotation about the carbon–nitrogen bond in a peptide linkage (see Section 24.8A) comes from ^{1}H NMR studies done with simple amides. For example, at room temperature the ^{1}H NMR spectrum of *N,N*-dimethylformamide, $(CH_3)_2NCHO$, shows a doublet at δ 2.80 (3H), a doublet at δ 2.95 (3H), and a multiplet at δ 8.05 (1H). When the spectrum is determined at lower magnetic field strength the doublets are found to have shifted so that the distance (in hertz) that separates one doublet from the other is smaller. When the temperature at which the spectrum is determined is raised, the doublets persist until a temperature of 111°C is reached; then the doublets coalesce to become a single signal. Explain in detail how these observations are consistent with the existence of a relatively large barrier to rotation about the carbon–nitrogen bond of DMF.

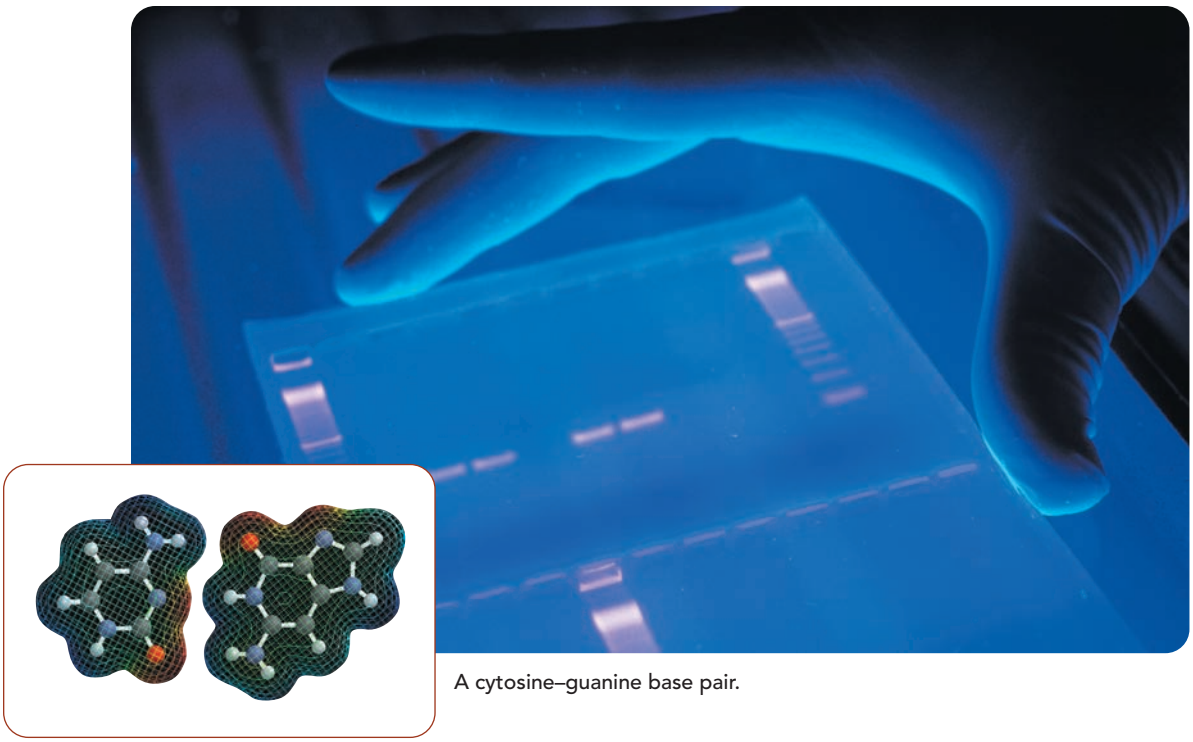
Learning Group Problems

1. The enzyme lysozyme and its mechanism are described in Section 24.10. Using the information presented there (and perhaps with additional information from a biochemistry textbook), prepare notes for a class presentation on the mechanism of lysozyme.

2. Chymotrypsin is a member of the serine protease class of enzymes. Its mechanism of action is described in Section 24.11. Using the information presented there (and perhaps supplemented by information from a biochemistry textbook), prepare notes for a class presentation on the mechanism of chymotrypsin. Consider especially the role of the "catalytic triad" with regard to acid–base catalysis and the relative propensity of various groups to act as nucleophiles or leaving groups.

25

Nucleic Acids and Protein Synthesis



A cytosine–guanine base pair.

Chemistry has long been called the central science—it is involved in every aspect of life. Much of what we have learned about chemistry is related to how things work, how diseases can be treated at the molecular level, and how materials we need in our daily lives can be improved or new ones created. Certainly not the least of chemistry's many applications, however, is an important dimension regarding work for global human rights and justice. As we are all too well aware, in many parts of the world there are situations where people have been separated from relatives because of the atrocious acts of war. Some scientists are tracing the family connections left after these grievous events using modern tools of chemistry. Laboratories such as those of M.-C. King (University of Washington) are attempting to help families bring closure when only remains of suspected relatives have been found and to reunite people in cases where victims have survived and they or their families are searching for familial ties.

The key to this work is DNA—the chemical fingerprint present in every tissue of every individual. Although the general structure of DNA is the same from one person to another, evidence for familial ties is present in the detailed sequence of each person's DNA. With the use of relatively simple chemistry—involving fluorescent dyes or radioactive isotopes, enzymes, gel electrophoresis, and a process called the polymerase chain reaction (PCR) that earned its inventor the 1993 Nobel Prize in Chemistry (Section 25.8)—it is now easy to synthesize millions of copies from a sample of DNA and to sequence it rapidly and conveniently. Application of these tools to comparison of DNA samples from victims and relatives provides hope that, at least in some cases, the gap between family members will be closed.

25.1 Introduction

Deoxyribonucleic acid (DNA) and **ribonucleic acid (RNA)** are molecules that carry genetic information in cells. DNA is the molecular archive of instructions for protein synthesis. RNA molecules transcribe and translate the information from DNA for the mechanics of protein synthesis. The storage of genetic information, its passage from generation to generation, and the use of genetic information to create the working parts of the cell all depend on the molecular structures of DNA and RNA. For these reasons, we shall focus our attention on the structures and properties of these **nucleic acids** and of their components, nucleotides and nucleosides.

DNA is a biological polymer composed of two molecular strands held together by hydrogen bonds. Its overall structure is that of a twisted ladder with a backbone of alternating sugar and phosphate units and rungs made of hydrogen-bonded pairs of heterocyclic amine bases (Fig. 25.1). DNA molecules are very long polymers. If the DNA from a single human cell were extracted and laid straight end-to-end, it would be roughly a meter long. To package DNA into the microscopic container of a cell's nucleus, however, it is supercoiled and bundled into the 23 pairs of chromosomes with which we are familiar from electron micrographs.

Four types of heterocyclic bases are involved in the rungs of the DNA ladder, and it is the sequence of these bases that carries the information for protein synthesis. Human DNA consists of approximately 3 billion base pairs. In an effort that marks a milestone in the history of science, a working draft of the sequence of the 3 billion base pairs in the human

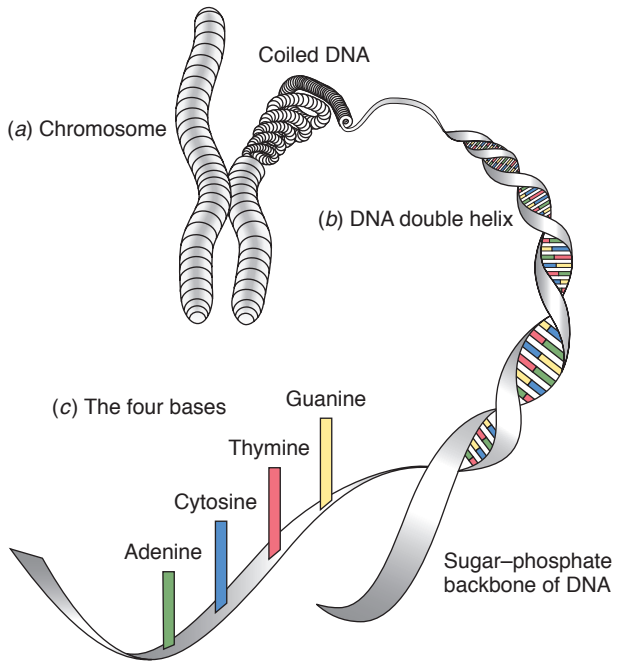


Figure 25.1 The basics of genetics. Each cell in the human body (except red blood cells) contains 23 pairs of chromosomes. Chromosomes are inherited: each parent contributes one chromosome per pair to their children. *(a)* Each chromosome is made up of a tightly coiled strand of DNA. The structure of DNA in its uncoiled state reveals *(b)* the familiar double-helix shape. If we picture DNA as a twisted ladder, the sides, made of sugar and phosphate molecules, are connected by *(c)* rungs made of heterocyclic amine bases. DNA has four, and only four, bases—adenine (A), thymine (T), guanine (G), and cytosine (C)—that form interlocking pairs. The order of the bases along the length of the ladder is called the DNA sequence. Within the overall sequence are genes, which encode the structure of proteins. (*Science and Technology Review*, November 1996, "The Human Genome Project," http://www..llnl.gov/str/Ashworth.html. Credit must be given to Linda Ashworth, the University of California, Lawrence Livermore National Laboratory, and the Department of Energy under whose auspices the work was performed, when this information or a reproduction of it is used.)

genome was announced in 2000. A final version was announced in 2003, the 50th anniversary of the structure determination of DNA by Watson and Crick.

- Each section of DNA that codes for a given protein is called a **gene**.
- The set of all genetic information coded by DNA in an organism is its **genome**.

There are approximately 30,000–35,000 genes in the human genome. The set of all proteins encoded within the genome of an organism and expressed at any given time is called its **proteome** (Section 24.14). Some scientists estimate there could be up to one million different proteins in the cells of our various tissues—a number much greater than the number of genes in the genome due to gene splicing during protein expression and post-translational protein modification.

Hopes are very high that, having sequenced the human genome, knowledge of it will bring increased identification of genes related to disease states (Fig. 25.2) and study of these genes and the proteins encoded by them will yield a myriad of benefits for human health and longevity. Determining the structure of all of the proteins encoded in the genome, learning their functions, and creating molecular therapeutics based on this rapidly expanding store of knowledge are some of the key research challenges that lie ahead.

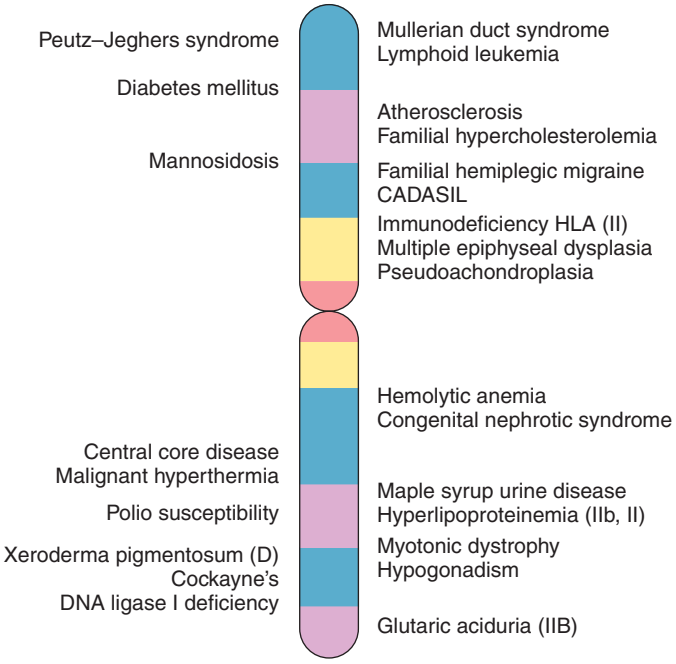


Figure 25.2 Schematic map of the location of genes for diseases on chromosome 19. (*From Dept. of Energy Joint Genome Institute Website (http://www.jgi.doe.gov/whoweare/llnl_jgi_decoding.htm#top). Credit to the University of California, Lawrence Livermore National Laboratory, and the Department of Energy under whose auspices the work was performed.*)

Let us begin with a study of the structures of nucleic acids. Each of their monomer units contains a cyclic amine base, a carbohydrate group, and a phosphate ester.

25.2 Nucleotides and Nucleosides

Mild degradations of nucleic acids yield monomeric units called **nucleotides**. A general formula for a nucleotide and the specific structure of one called adenylic acid are shown in Fig. 25.3.

Complete hydrolysis of a nucleotide furnishes:

1. A heterocyclic base from either the purine or pyrimidine family.
2. A five-carbon monosaccharide that is either D-ribose or 2-deoxy-D-ribose.
3. A phosphate ion.

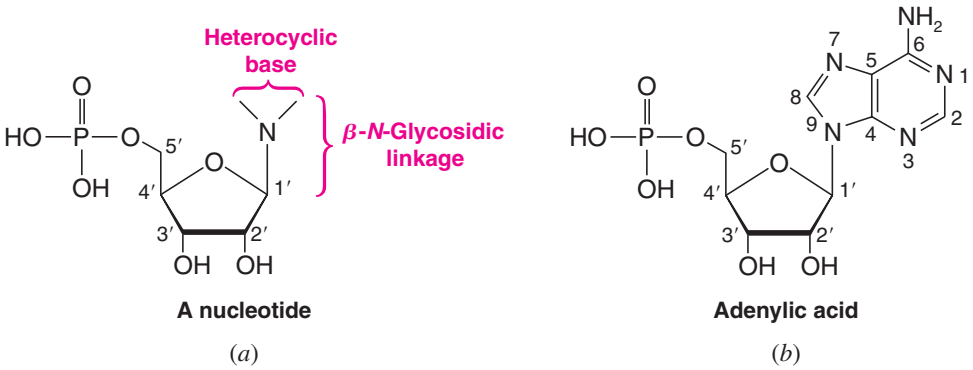
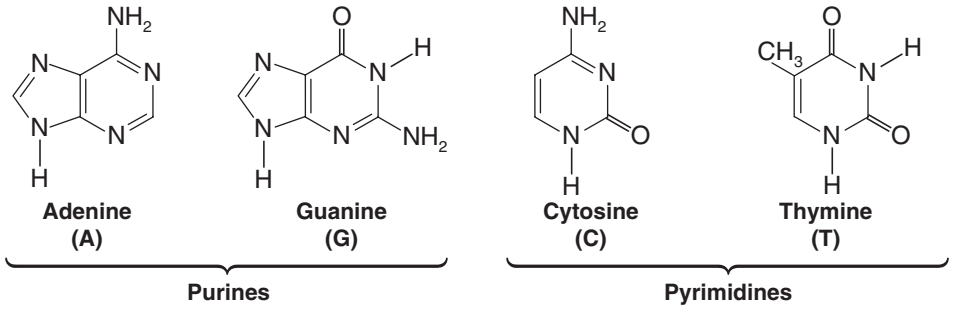


Figure 25.3 *(a)* General structure of a nucleotide obtained from RNA. The heterocyclic base is a purine or pyrimidine. *In nucleotides obtained from DNA, the sugar component is 2′-deoxy-D-ribose; that is, the —OH at position 2′ is replaced by —H.* The phosphate group of the nucleotide is shown attached at C5′; it may instead be attached at C3′. In DNA and RNA a phosphodiester linkage joins C5′ of one nucleotide to C3′ of another. The heterocyclic base is always attached through a β-*N*-glycosidic linkage at C1′. *(b)* Adenylic acid, a typical nucleotide.

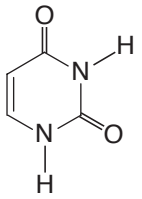
The central portion of the nucleotide is the monosaccharide, and it is always present as a five-membered ring, that is, as a furanoside. The heterocyclic base of a nucleotide is attached through an *N*-glycosidic linkage to C1′ of the ribose or deoxyribose unit, and this linkage is always β. The phosphate group of a nucleotide is present as a phosphate ester and may be attached at C5′ or C3′. (In nucleotides, the carbon atoms of the monosaccharide portion are designated with primed numbers, i.e., 1′, 2′, 3′, etc.)

Removal of the phosphate group of a nucleotide converts it to a compound known as a **nucleoside** (Section 22.15A). The nucleosides that can be obtained from DNA all contain 2-deoxy-D-ribose as their sugar component and one of four heterocyclic bases: adenine, guanine, cytosine, or thymine:



The nucleosides obtained from RNA contain D-ribose as their sugar component and adenine, guanine, cytosine, or uracil as their heterocyclic base.

Uracil replaces thymine in an RNA nucleoside (or nucleotide). (Some nucleosides obtained from specialized forms of RNA may also contain other, but similar, purines and pyrimidines.)



The heterocyclic bases obtained from nucleosides are capable of existing in more than one tautomeric form. The forms that we have shown are the predominant forms that the bases assume when they are present in nucleic acids.

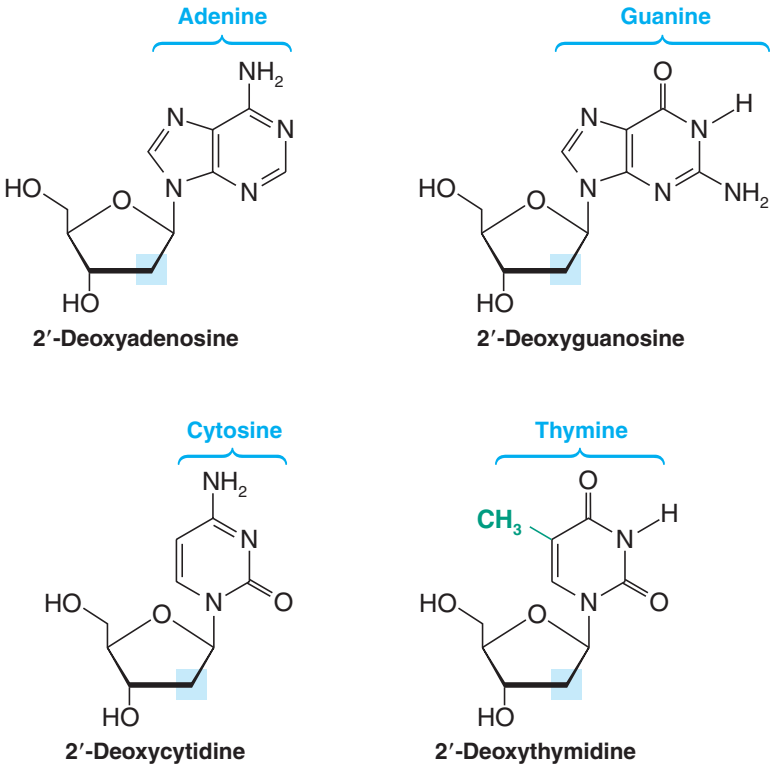


Figure 25.4 Nucleosides that can be obtained from DNA. DNA is 2′-deoxy at the position where the blue shaded box is shown. RNA (see Fig 25.5) has hydroxyl groups at that location. RNA has a hydrogen where there is a methyl group in thymine, which in RNA makes the base uracil (and the nucleoside uridine).

The names and structures of the nucleosides found in DNA are shown in Fig. 25.4; those found in RNA are given in Fig. 25.5.

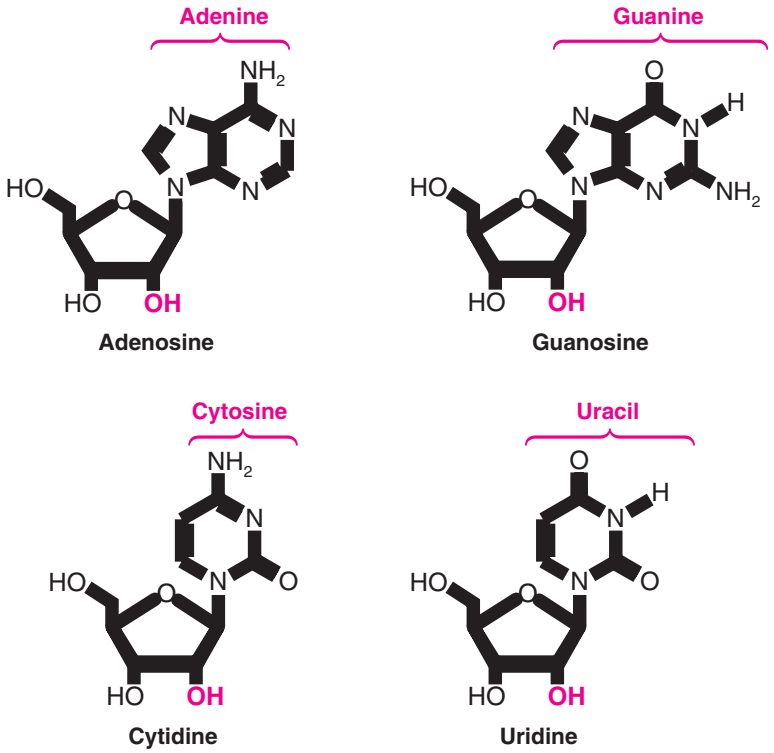


Figure 25.5 Nucleosides that can be obtained from RNA. DNA (see Fig 25.4) has hydrogen atoms where the red hydroxyl groups of ribose are shown (DNA is 2′-deoxy with respect to its ribose moiety).

Review Problem 25.1 Write the structures of other tautomeric forms of adenine, guanine, cytosine, thymine, and uracil.

Review Problem 25.2 The nucleosides shown in Figs. 25.4 and 25.5 are stable in dilute base. In dilute acid, however, they undergo rapid hydrolysis yielding a sugar (deoxyribose or ribose) and a heterocyclic base.

(a) What structural feature of the nucleoside accounts for this behavior?

(b) Propose a reasonable mechanism for the hydrolysis.

Nucleotides are named in several ways. Adenylic acid (Fig. 25.3), for example, is usually called AMP, for adenosine monophosphate. The position of the phosphate group is sometimes explicitly noted by use of the names adenosine 5′-monophosphate or 5′-adenylic acid. Uridylic acid is usually called UMP, for uridine monophosphate, although it can also be called uridine 5′-monophosphate or 5′-uridylic acid. If a nucleotide is present as a diphosphate or triphosphate, the names are adjusted accordingly, such as ADP for adenosine diphosphate or GTP for guanosine triphosphate.

Nucleosides and nucleotides are found in places other than as part of the structure of DNA and RNA. We have seen, for example, that adenosine units are part of the structures of two important coenzymes, NADH and coenzyme A. The 5′-triphosphate of adenosine is, of course, the important energy source, ATP (Section 22.1B). The compound called 3′,5′-cyclic adenylic acid (or cyclic AMP) (Fig. 25.6) is an important regulator of hormone activity. Cells synthesize this compound from ATP through the action of an enzyme, *adenylate cyclase*. In the laboratory, 3′,5′-cyclic adenylic acid can be prepared through dehydration of 5′-adenylic acid with dicyclohexylcarbodiimide.

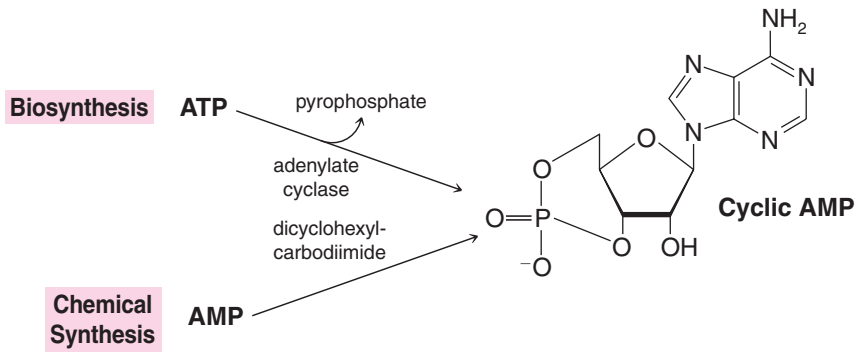


Figure 25.6 3′,5′-Cyclic adenylic acid (cyclic AMP) and its biosynthesis and laboratory synthesis.

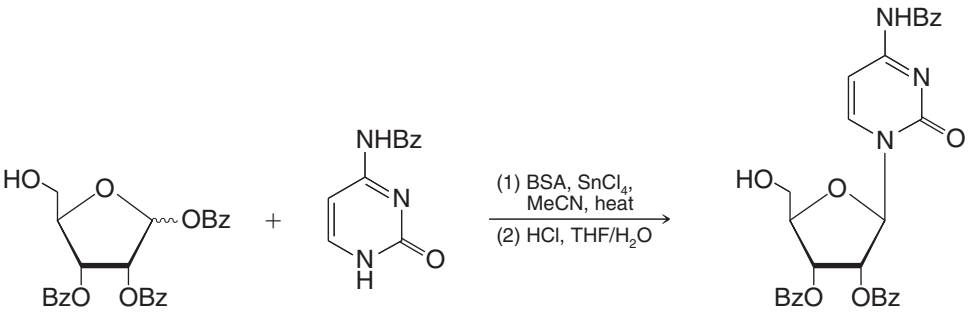
Solved Problem 25.1

When 3′,5′-cyclic adenylic acid is treated with aqueous sodium hydroxide, the major product that is obtained is 3′-adenylic acid (adenosine 3′-phosphate) rather than 5′-adenylic acid. Suggest a mechanism that explains the course of this reaction.

STRATEGY AND ANSWER The reaction appears to take place through an S_N2 mechanism. Attack occurs preferentially at the primary 5′-carbon atom rather than at the secondary 3′-carbon atom due to the difference in steric hindrance.

25.3 Laboratory Synthesis of Nucleosides and Nucleotides

A variety of methods have been developed for the chemical synthesis of nucleosides from the constituent sugars and bases or their precursors. The following is an example of a *silyl–Hilbert–Johnson nucleosidation*, where a benzoyl protected sugar (D-ribose) reacts in the presence of tin(IV) chloride with an *N*-benzoyl protected base (cytidine) that is protected further by *in situ* silylation.* The trimethylsilyl protecting groups for the base are introduced using *N,O*-bis(trimethylsilyl)acetamide (BSA) and they are removed with aqueous acid in the second step. The result is a protected form of the nucleoside cytosine, from which the benzoyl groups can be removed with ease using a base:

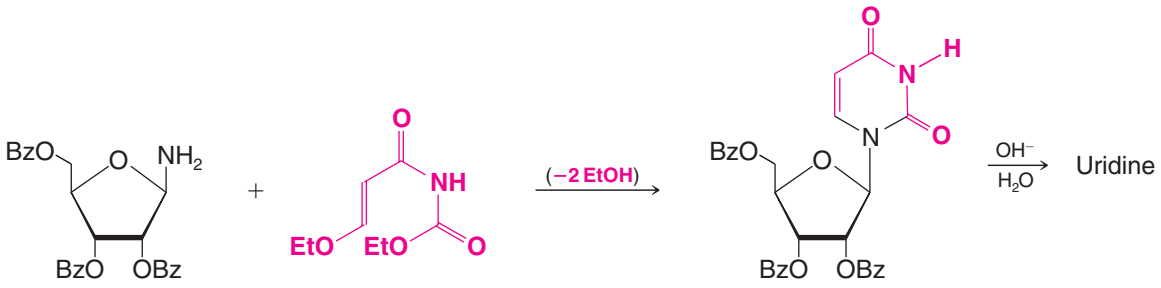


D-Ribose (Bz protected) **N^4-Benzoylcytosine** **Cytidine (Bz protected)**

Bz = C_6H_5CO (benzoyl)

BSA = $CH_3C[{=}NSi(CH_3)_3]OSi(CH_3)_3$, [*N,O*-bis(trimethylsilyl)acetamide, a reagent for trimethylsilylation of nucleophilic oxygen and nitrogen atoms]

Another technique involves formation of the heterocyclic base on a protected ribosylamine derivative:



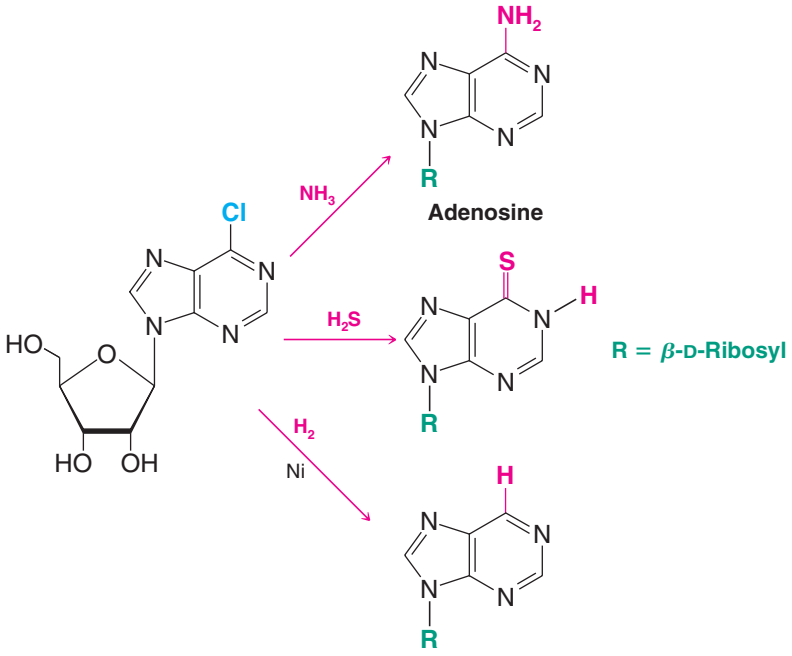
2,3,5-Tri-*O*-benzoyl-β-D-ribofuranosylamine **β-Ethoxy-*N*-ethoxycarbonylacrylamide**

Basing your answer on reactions that you have seen before, propose a likely mechanism for the condensation reaction in the first step of the preceding uridine synthesis.

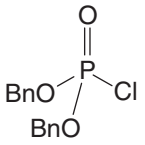
Review Problem 25.3

*These conditions were applied using L-ribose in a synthesis of the unnatural enantiomer of RNA (Pitsch, S. an efficient synthesis of enantiomeric ribonucleic acids from D-glucose. *Helv. Chim. Acta* **1997**, *80*, 2286–2314). The protected enantiomeric cytidine was produced in 94% yield by the above reaction. After adjusting protecting groups, solid-phase oligonucleotide synthesis methods (Section 25.7) were used with this compound and the other three nucleotide monomers (also derived from L-ribose) for preparation of the unnatural RNA enantiomer. See also Vorbrüggen, H.; Ruh-Pohlenz, C., *Handbook of Nucleoside Synthesis*; Wiley: Hoboken, NJ, 2001.

Still a third technique involves the synthesis of a nucleoside with a substituent in the heterocyclic ring that can be replaced with other groups. This method has been used extensively to synthesize unusual nucleosides that do not necessarily occur naturally. The following example makes use of a 6-chloropurine derivative obtained from the appropriate ribofuranosyl chloride and chloromercuripurine:

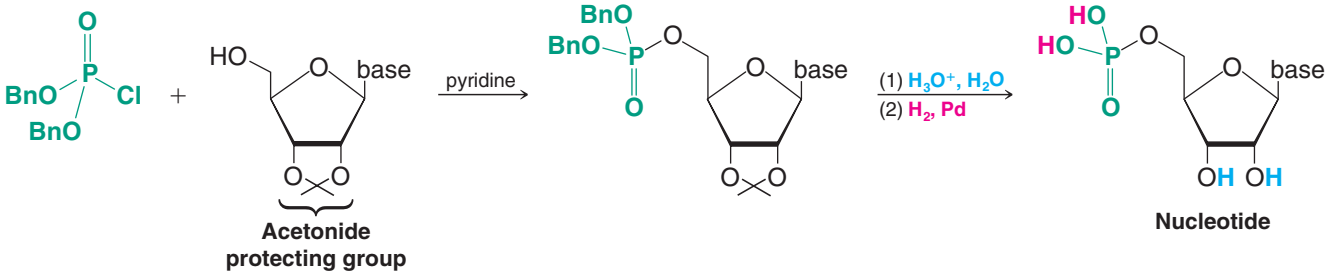


Numerous phosphorylating agents have been used to convert nucleosides to nucleotides. One of the most useful is dibenzyl phosphochloridate:



Bn = $C_6H_5CH_2$ **(benzyl)**

Specific phosphorylation of the 5′-OH can be achieved if the 2′- and 3′-OH groups of the nucleoside are protected by an acetonide group (see the following):



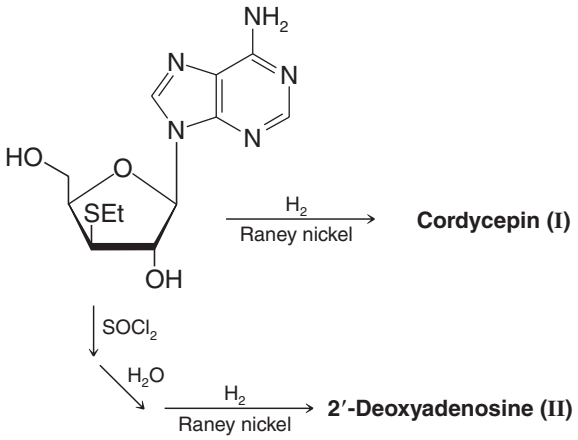
Mild acid-catalyzed hydrolysis removes the acetonide group, and hydrogenolysis cleaves the benzyl phosphate bonds.

Review Problem 25.4

(a) What kind of linkage is involved in the acetonide group of the protected nucleoside, and why is it susceptible to mild acid-catalyzed hydrolysis? **(b)** How might such a protecting group be installed?

Review Problem 25.5

The following reaction scheme is from a synthesis of cordycepin (a nucleoside antibiotic) and the first synthesis of 2′-deoxyadenosine (reported in 1958 by C. D. Anderson, L. Goodman, and B. R. Baker, Stanford Research Institute):

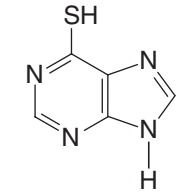


(a) What is the structure of cordycepin? (**I** and **II** are isomers.)

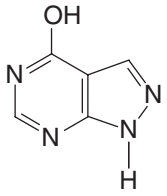
(b) Propose a mechanism that explains the formation of **II**.

25.3A Medical Applications

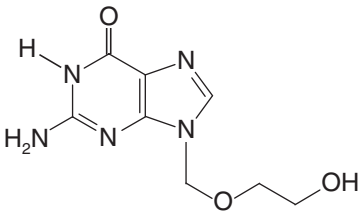
In the early 1950s, Gertrude Elion and George Hitchings (of the Wellcome Research Laboratories) discovered that 6-mercaptopurine had antitumor and antileukemic properties. This discovery led to the development of other purine derivatives and related compounds, including nucleosides, of considerable medical importance. Three examples are the following:



6-Mercaptopurine



Allopurinol



Acyclovir

Elion and Hitchings shared the 1988 Nobel Prize in Physiology or Medicine for their work in the development of chemotherapeutic agents derived from purines.

6-Mercaptopurine is used in combination with other chemotherapeutic agents to treat acute leukemia in children, and almost 80% of the children treated are now cured. Allopurinol, another purine derivative, is a standard therapy for the treatment of gout. Acyclovir, a nucleoside that lacks two carbon atoms of its ribose ring, is highly effective in treating diseases caused by certain herpes viruses, including *herpes simplex* type 1 (fever blisters), type 2 (genital herpes), and varicella-zoster (shingles).

25.4 Deoxyribonucleic Acid: DNA

25.4A Primary Structure

Nucleotides bear the same relation to a nucleic acid that amino acids do to a protein; they are its monomeric units. The connecting links in proteins are amide groups; in nucleic acids they are phosphate ester linkages. Phosphate esters link the 3′-OH of one ribose (or deoxyribose) with the 5′-OH of another. This makes the nucleic acid a long unbranched chain with

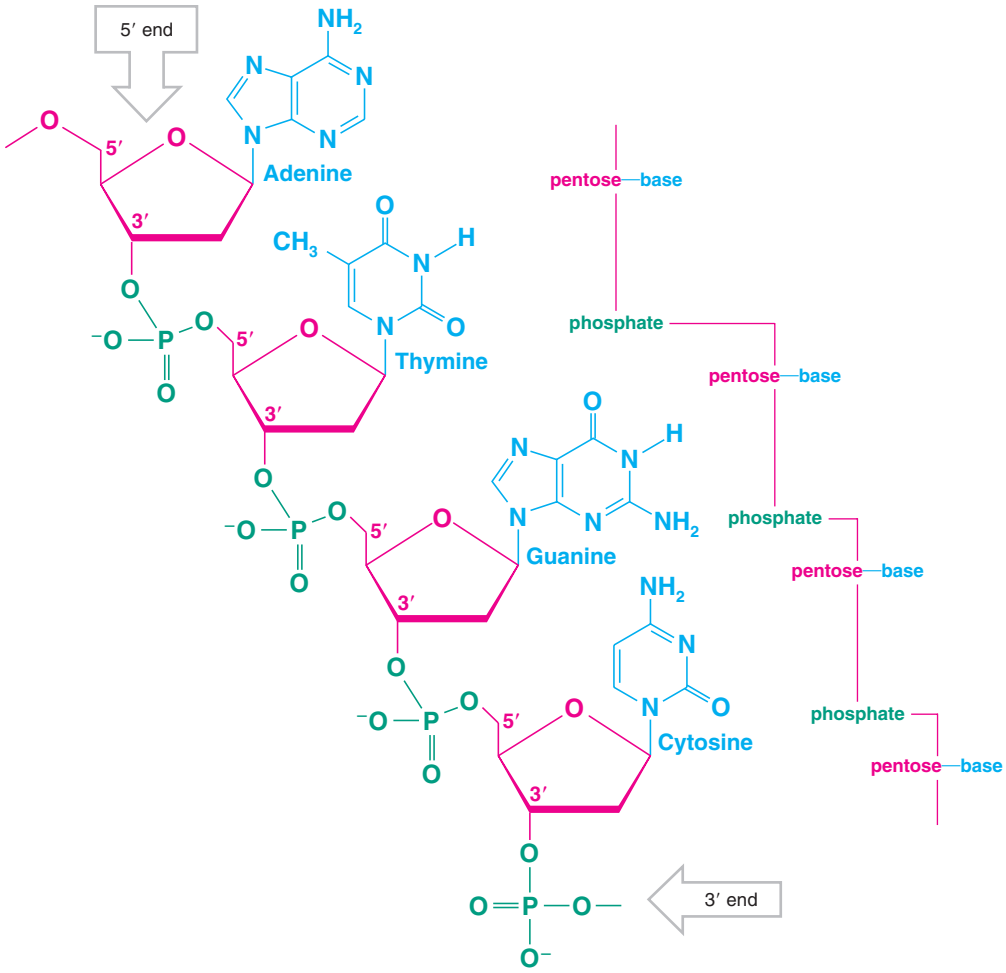


Figure 25.7 A segment of one DNA chain showing how phosphate ester groups link the 3′- and 5′-OH groups of deoxyribose units. RNA has a similar structure with two exceptions: A hydroxyl replaces a hydrogen atom at the 2′ position of each ribose unit and uracil replaces thymine.

a "backbone" of sugar and phosphate units with heterocyclic bases protruding from the chain at regular intervals (Fig. 25.7). We would indicate the direction of the bases in Fig. 25.7 in the following way:

$$5' \leftarrow A—T—G—C \rightarrow 3'$$

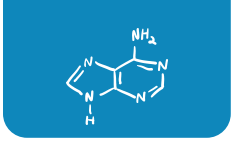
It is, as we shall see, the **base sequence** along the chain of DNA that contains the encoded genetic information. The sequence of bases can be determined using enzymatic methods and chromatography (Section 25.6).

25.4B Secondary Structure

"I cannot help wondering whether some day an enthusiastic scientist will christen his newborn twins Adenine and Thymine."
F. H. C. Crick*

It was the now-classic proposal of James Watson and Francis Crick (made in 1953 and verified shortly thereafter through the X-ray analysis by Maurice Wilkins) that gave a model for the secondary structure of DNA. This work earned Crick, Watson, and Wilkins the 1962 Nobel Prize in Physiology or Medicine. Many believe that Rosalind Franklin, whose X-ray data was also key to solving the structure of DNA, should have shared the Nobel prize, but her death from cancer in 1958 precluded it. The secondary structure of DNA is especially important because it enables us to understand how genetic information is preserved, how it can be passed on during the process of cell division, and how it can be transcribed to provide a template for protein synthesis.

*Taken from Crick, F. H. C., The structure of the hereditary material. *Sci. Am.* **1954**, *191*(10), 20, 54–61.



Of prime importance to Watson and Crick's proposal was an earlier observation (made in the late 1940s) by Erwin Chargaff that certain regularities can be seen in the percentages of heterocyclic bases obtained from the DNA of a variety of species. Table 25.1 gives results that are typical of those that can be obtained.

TABLE 25.1 DNA Composition of Various Species

Species	Base Proportions (mol %)							
	G	A	C	T	$\frac{G+A}{C+T}$	$\frac{A+T}{G+C}$	$\frac{A}{T}$	$\frac{G}{C}$
Sarcina lutea	37.1	13.4	37.1	12.4	1.02	0.35	1.08	1.00
Escherichia coli K12	24.9	26.0	25.2	23.9	1.08	1.00	1.09	0.99
Wheat germ	22.7	27.3	22.8[a]	27.1	1.00	1.19	1.01	1.00
Bovine thymus	21.5	28.2	22.5[a]	27.8	0.96	1.27	1.01	0.96
Staphylococcus aureus	21.0	30.8	19.0	29.2	1.11	1.50	1.05	1.11
Human thymus	19.9	30.9	19.8	29.4	1.01	1.52	1.05	1.01
Human liver	19.5	30.3	19.9	30.3	0.98	1.54	1.00	0.98

[a]Cytosine + methylcytosine.

Source: Smith, E. L.; Hill, R. L.; Lehman, I. R.; Lefkowitz, R. J.; Handler, P.; and White, A. *Principles of Biochemistry: General Aspects*, 7th ed. McGraw-Hill: New York, 1983; p. 132. Copyright © 1983 by McGraw-Hill, Inc. Reproduced with permission of McGraw-Hill Companies.

Chargaff pointed out that for all species examined:

1. The total mole percentage of purines is approximately equal to that of the pyrimidines, that is, (%G + %A)/(%C + %T) ≅ 1.
2. The mole percentage of adenine is nearly equal to that of thymine (i.e., %A/%T ≅ 1), and the mole percentage of guanine is nearly equal to that of cytosine (i.e., %G/%C ≅ 1).

Chargaff also noted that the ratio which varies from species to species is the ratio (%A + %T)/(%G + %C). He noted, moreover, that whereas this ratio is characteristic of the DNA of a given species, it is the same for DNA obtained from different tissues of the same animal and does not vary appreciably with the age or conditions of growth of individual organisms within the same species.

Watson and Crick also had X-ray data that gave them the bond lengths and angles of the purine and pyrimidine rings of model compounds. In addition, they had data from Franklin and Wilkins that indicated a repeat distance of 34 Å in DNA.

Reasoning from these data, Watson and Crick proposed a double helix as a model for the secondary structure of DNA. According to this model, two nucleic acid chains are held together by hydrogen bonds between base pairs on opposite strands. This double chain is wound into a helix with both chains sharing the same axis. The base pairs are on the inside of the helix, and the sugar–phosphate backbone is on the outside (Fig. 25.8). The pitch of the helix is such that 10 successive nucleotide pairs give rise to one complete turn in 34 Å (the repeat distance). The exterior width of the spiral is about 20 Å, and the internal distance between 1′ positions of ribose units on opposite chains is about 11 Å.

Using molecular-scale models, Watson and Crick observed that the internal distance of the double helix is such that it allows only a purine–pyrimidine type of hydrogen bonding between base pairs. Purine–purine base pairs do not occur because they would be too large to fit, and pyrimidine–pyrimidine base pairs do not occur because they would be too far apart to form effective hydrogen bonds.

Helpful Hint

The use of models was critical to Watson and Crick in their Nobel prize–winning work on the three-dimensional structure of DNA.

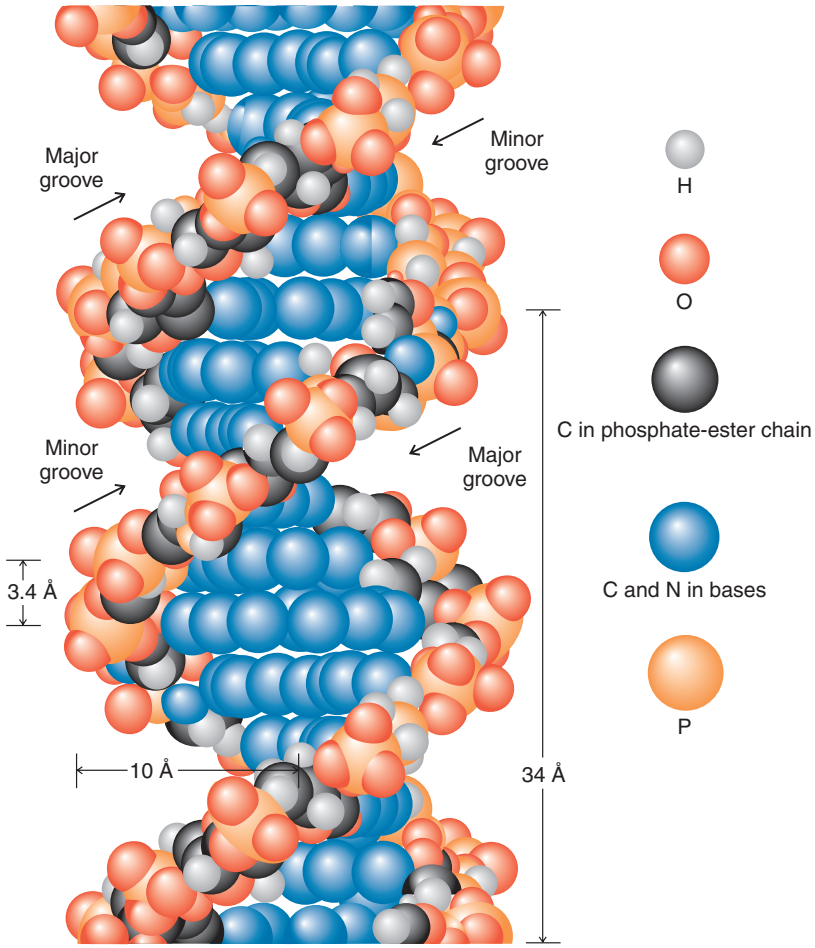
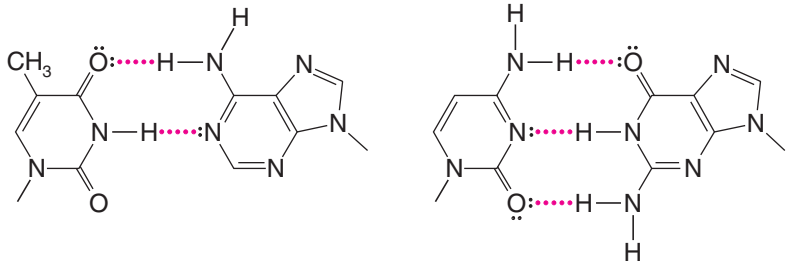


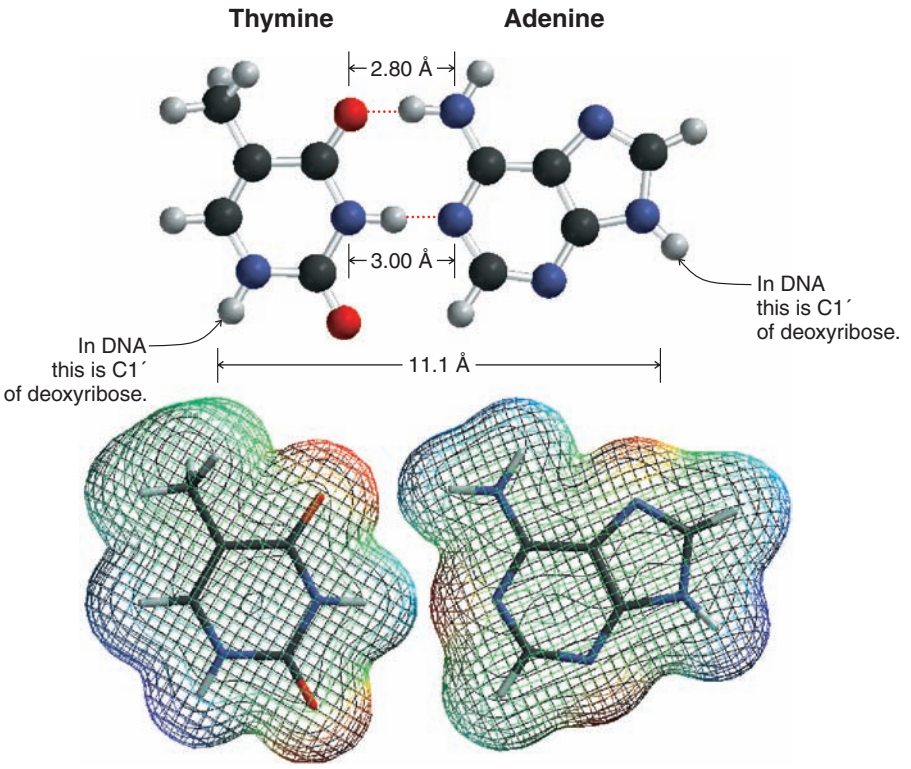
Figure 25.8 A molecular model of a portion of the DNA double helix. (Reprinted with permission of The McGraw-Hill Companies from Neal, L., *Chemistry and Biochemistry: A Comprehensive Introduction*, © 1971.)

Watson and Crick went one crucial step further in their proposal. Assuming that the oxygen-containing heterocyclic bases existed in keto forms, they argued that base pairing through hydrogen bonds can occur in only a specific way: adenine (A) with thymine (T) and cytosine (C) with guanine (G). Dimensions of the pairs and electrostatic potential maps for the individual bases are shown in Fig. 25.9.

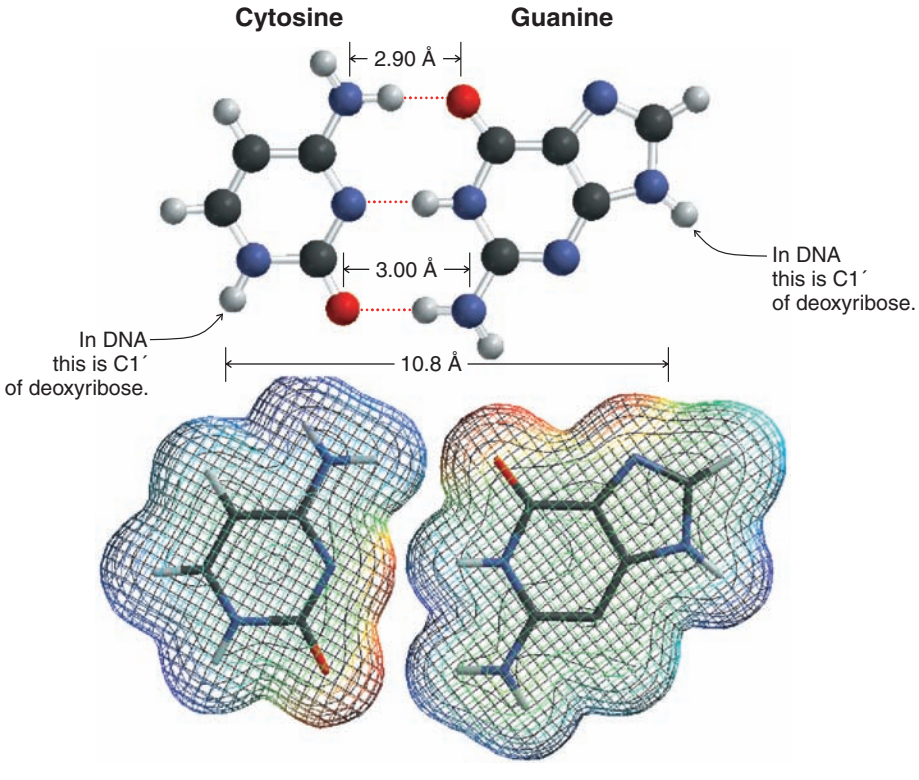
Adenine pairs with thymine **Guanine pairs with cytosine**



Specific base pairing of this kind is consistent with Chargaff's finding that %A/%T ≅ 1 and %G/%C ≅ 1.



(*a*)



(*b*)

Figure 25.9 Base pairing of adenine with thymine *(a)* and cytosine with guanine *(b)*. The dimensions of the thymine–adenine and cytosine–guanine hydrogen-bonded pairs are such that they allow the formation of strong hydrogen bonds and also allow the base pairs to fit inside the two phosphate–ribose chains of the double helix. (Reprinted from Archives of *Biochemistry and Biophysics*, **65**, Pauling, I., Corey, R., p. 164–181, 1956. Copyright 1956, with permission from Elsevier.) Electrostatic potential maps calculated for the individual bases show the complementary distribution of charges that leads to hydrogen bonding.

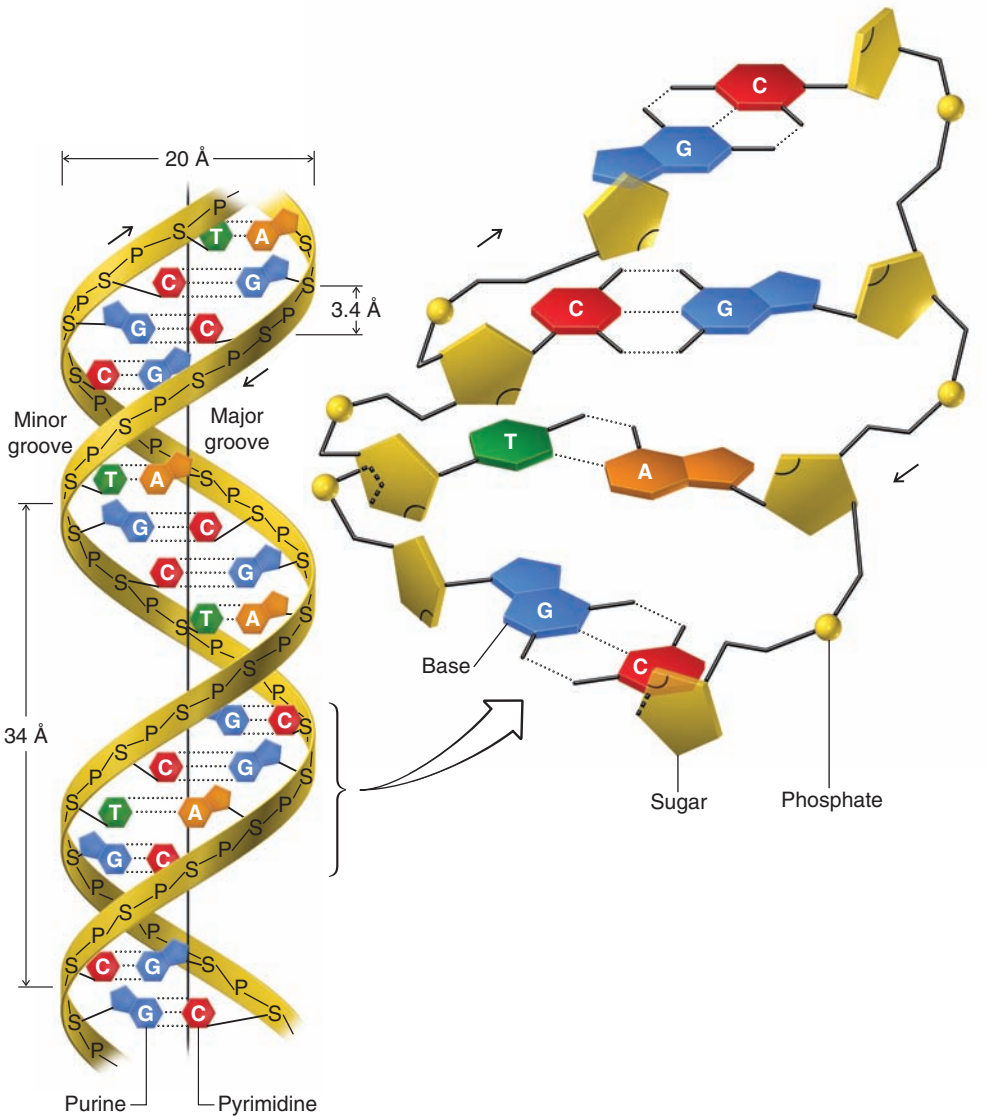


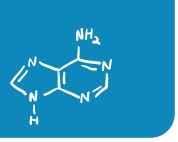
Figure 25.10 Diagram of the DNA double helix showing complementary base pairing. The arrows indicate the 3′ → 5′ direction.

Specific base pairing also means that the two chains of DNA are complementary. Wherever adenine appears in one chain, thymine must appear opposite it in the other; wherever cytosine appears in one chain, guanine must appear in the other (Fig. 25.10).

Notice that while the sugar–phosphate backbone of DNA is completely regular, the sequence of heterocyclic base pairs along the backbone can assume many different permutations. This is important because it is the precise sequence of base pairs that carries the genetic information. Notice, too, that one chain of the double strand is the complement of the other. If one knows the sequence of bases along one chain, one can write down the sequence along the other, because A always pairs with T and G always pairs with C. It is this complementarity of the two strands that explains how a DNA molecule replicates itself at the time of cell division and thereby passes on the genetic information to each of the two daughter cells.

25.4C Replication of DNA

Just prior to cell division the double strand of DNA begins to unwind. Complementary strands are formed along each chain (Fig. 25.11). Each chain acts, in effect, as a template



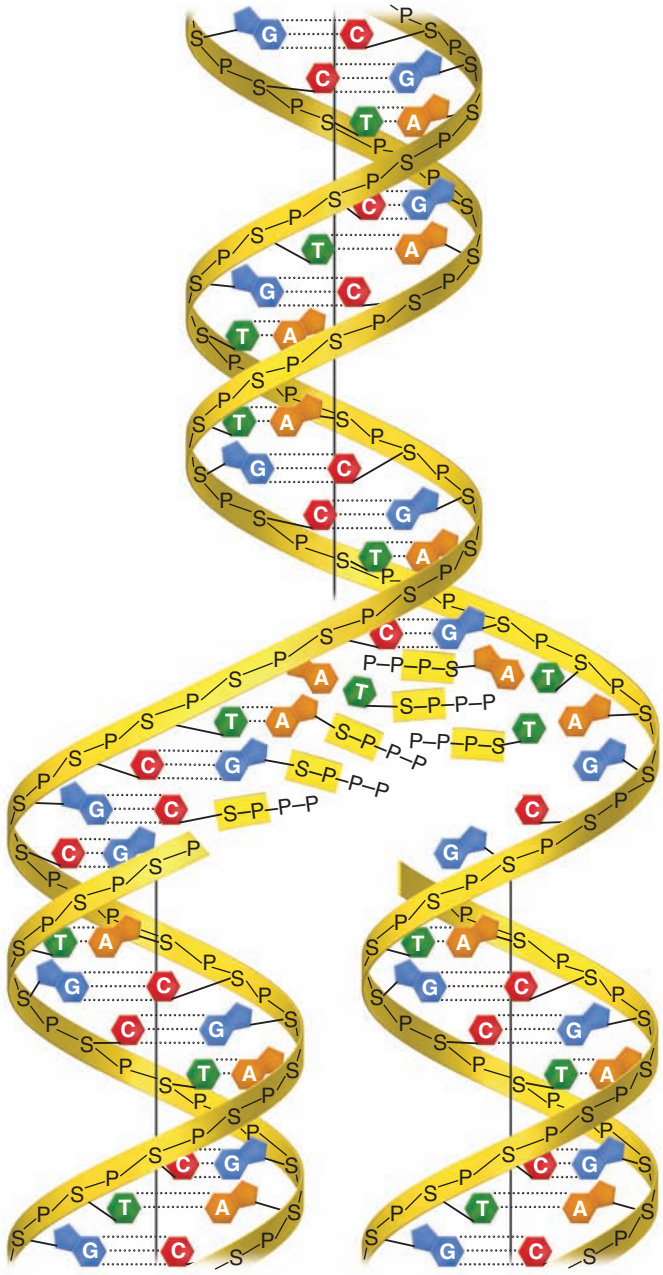


Figure 25.11 Replication of DNA. The double strand unwinds from one end and complementary strands are formed along each chain.

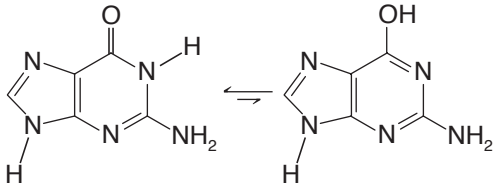
for the formation of its complement. When unwinding and **replication** are complete, there are two identical DNA molecules where only one had existed before. These two molecules can then be passed on, one to each daughter cell.

Review Problem 25.6

(a) There are approximately 3 billion base pairs in the DNA of a single human cell. Assuming that this DNA exists as a double helix, calculate the length of all the DNA contained in a human cell. **(b)** The weight of DNA in a single human cell is 6×10^{-12} g. Assuming that Earth's population is about 6.5 billion, we can conclude that all of the genetic information that gave rise to all human beings now alive was once contained in the DNA of a corresponding number of fertilized ova. What is the total weight of DNA in this many ova? (The volume that this DNA would occupy is approximately that of a raindrop, yet if the individual molecules were laid end-to-end, they would stretch to the moon and back almost eight times.)

Review Problem 25.7

(a) The most stable tautomeric form of guanine is the lactam form (or cyclic amide, see Section 17.8I). This is the form normally present in DNA, and, as we have seen, it pairs specifically with cytosine. If guanine tautomerizes (see Section 18.2) to the lactim form, it pairs with thymine instead. Write structural formulas showing the hydrogen bonds in this abnormal base pair.

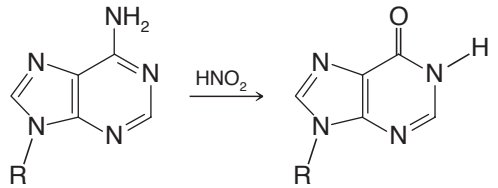


Lactam form of guanine **Lactim form of guanine**

(b) Improper base pairings that result from tautomerizations occurring during the process of DNA replication have been suggested as a source of spontaneous mutations. We saw in part (a) that if a tautomerization of guanine occurred at the proper moment, it could lead to the introduction of thymine (instead of cytosine) into its complementary DNA chain. What error would this new DNA chain introduce into *its* complementary strand during the next replication even if no further tautomerizations take place?

Review Problem 25.8

Mutations can also be caused chemically, and nitrous acid is one of the most potent chemical **mutagens**. One explanation that has been suggested for the mutagenic effect of nitrous acid is the deamination reactions that it causes with purines and pyrimidines bearing amino groups. When, for example, an adenine-containing nucleotide is treated with nitrous acid, it is converted to a hypoxanthine derivative:



Adenine nucleotide **Hypoxanthine nucleotide**

(a) Basing your answer on reactions you have seen before, what are likely intermediates in the adenine ⟶ hypoxanthine interconversion? **(b)** Adenine normally pairs with thymine in DNA, but hypoxanthine pairs with cytosine. Show the hydrogen bonds of a hypoxanthine–cytosine base pair. **(c)** Show what errors an adenine ⟶ hypoxanthine interconversion would generate in DNA through two replications.

25.5 *RNA and Protein Synthesis*

Soon after the Watson–Crick hypothesis was published, scientists began to extend it to yield what Crick called "the central dogma of molecular genetics." This dogma stated that genetic information flows as follows:

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{protein}$$

The synthesis of protein is, of course, all important to a cell's function because proteins (as enzymes) catalyze its reactions. Even the very primitive cells of bacteria require as many as 3000 different enzymes. This means that the DNA molecules of these cells must contain a corresponding number of genes to direct the synthesis of these proteins. A **gene** is that segment of the DNA molecule that contains the information necessary to direct the synthesis of one protein (or one polypeptide).

There are viruses, called retroviruses, in which information flows from RNA to DNA. The virus that causes AIDS is a retrovirus.

DNA is found primarily in the nucleus of eukaryotic cells. Protein synthesis takes place primarily in that part of the cell called the *cytoplasm*. Protein synthesis requires that two major processes take place; the first occurs in the cell nucleus, the second in the cytoplasm. The first is **transcription**, a process in which the genetic message is transcribed onto a form of RNA called messenger RNA (mRNA). The second process involves two other forms of RNA, called ribosomal RNA (rRNA) and transfer RNA (tRNA).

25.5A Messenger RNA Synthesis—Transcription

The events leading to protein synthesis begin in the cell nucleus with the synthesis of mRNA. Part of the DNA double helix unwinds sufficiently to expose on a single chain a portion corresponding to at least one gene. Ribonucleotides, present in the cell nucleus, assemble along the exposed DNA chain by pairing with the bases of DNA. The pairing patterns are the same as those in DNA with the exception that in RNA uracil replaces thymine. The ribonucleotide units of mRNA are joined into a chain by an enzyme called *RNA polymerase*. This process is illustrated in Fig. 25.12.

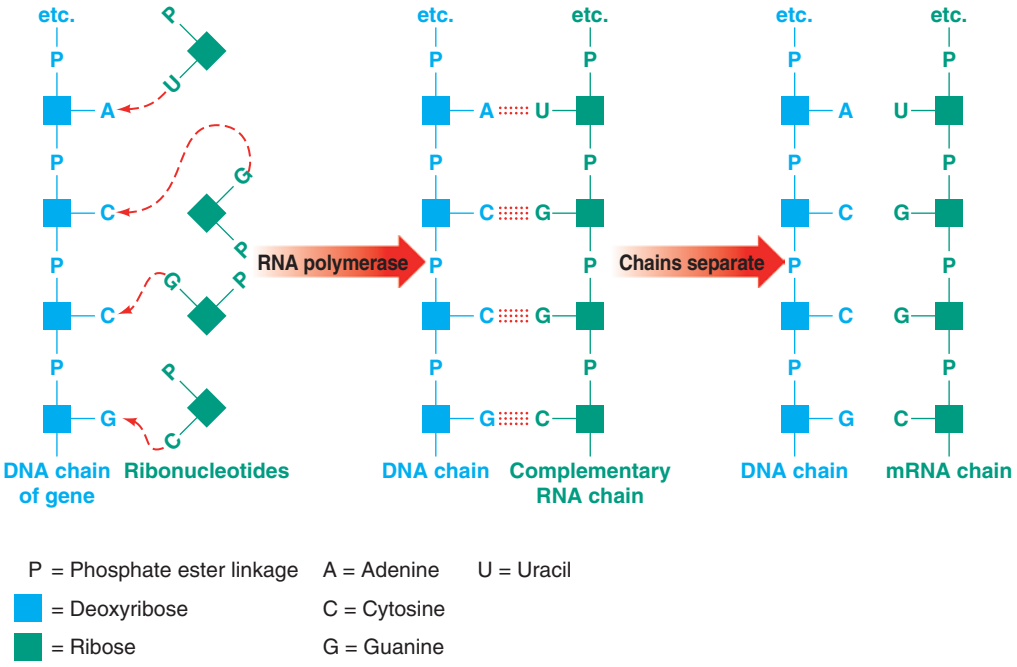


Figure 25.12 Transcription of the genetic code from DNA to mRNA.

Review Problem 25.9

Write structural formulas showing how the keto form of uracil (Section 25.2) in mRNA can pair with adenine in DNA through hydrogen bond formation.

Most eukaryotic genes contain segments of DNA that are not actually used when a protein is expressed, even though they are transcribed into the initial mRNA. These segments are called **introns**, or intervening sequences. The segments of DNA within a gene that are expressed are called **exons**, or expressed sequences. Each gene usually contains a number of introns and exons. After the mRNA is transcribed from DNA, the introns in the mRNA are removed and the exons are spliced together.

After mRNA has been synthesized and processed in the cell nucleus to remove the introns, it migrates into the cytoplasm where, as we shall see, it acts as a template for protein synthesis.

25.5B Ribosomes—rRNA

Protein synthesis is catalyzed by ribosomes in the cytoplasm. Ribosomes (Fig. 25.13) are ribonucleoproteins, comprised of approximately two-thirds RNA and one-third protein. They have a very high molecular weight (about 2.6×10^6). The RNA component is present in two subunits, called the 50S and 30S subunits (classified according to their sedimentation behavior during ultracentrifugation*). The 50S subunit is roughly twice the molecular weight of the 30S subunit. Binding of RNA with mRNA is mediated by the 30S subunit. The 50S subunit carries the catalytic activity for translation that joins one amino acid by an amide bond to the next. In addition to the rRNA subunits there are approximately 30–35 proteins tightly bound to the ribosome, the entire structure resembling an exquisite three-dimensional jigsaw puzzle of RNA and protein. The mechanism for ribosome-catalyzed amide bond formation is discussed below.

Ribosomes, as reaction catalysts, are most appropriately classified as **ribozymes** rather than enzymes, because it is RNA that catalyzes the peptide bond formation during protein synthesis and not the protein subunits of the ribosome. The mechanism for peptide bond formation catalyzed by the 50S ribosome subunit (Fig. 25.14), proposed by Moore and coworkers based on X-ray crystal structures, suggests that attack by the α-amino group is facilitated by acid–base catalysis involving nucleotide residues along the 50S ribosome subunit chain, specifically a nearby adenine group. Full or partial removal of a proton from the α-amino group of the amino acid by N3 of the adenine group imparts greater nucleophilicity to the amino nitrogen, facilitating its attack on the acyl carbon of the adjacent peptide–tRNA moiety. A tetrahedral intermediate is formed, which collapses to form the new amide bond with release of the tRNA that had been joined to the peptide. Other moieties in the 50S ribosome subunit are believed to help stabilize the transfer of charge that occurs as N3 of the adenyl group accepts the proton from the α-amino group of the new amino acid (see Problem 25.16).

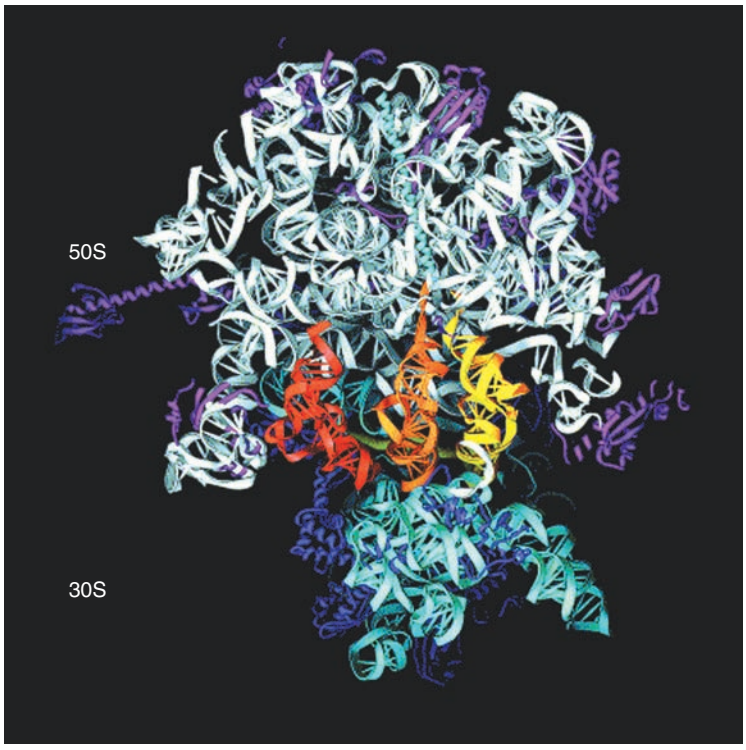
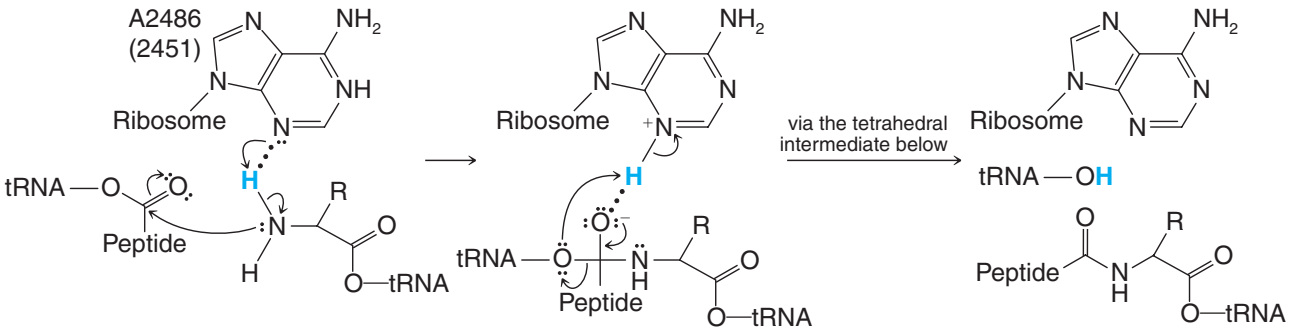


Figure 25.13 Structure of the *Thermus thermophilus* ribosome showing the 50S and 30S subunits and three bound transfer RNAs. The yellow tRNA is at the A site, which would bear the new amino acid to be added to the peptide. The light orange tRNA is at the P site, which would be the tRNA that bears the growing peptide. The red tRNA is at the E site, which is the "empty" tRNA after it has transferred the peptide chain to the new amino acid. (*Courtesy of Harry Noller, University of California, Santa Cruz.*)

*S stands for svedberg unit; it is used in describing the behavior of proteins in an ultracentrifuge.



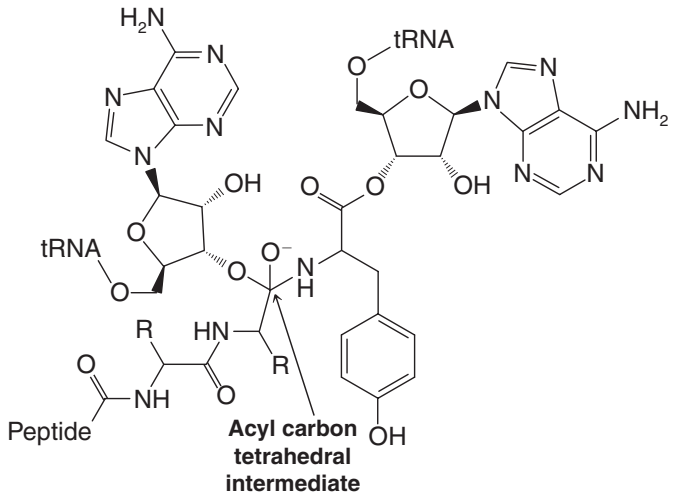


Figure 25.14 A mechanism for peptide bond formation catalyzed by the 50S subunit of the ribosome (as proposed by Moore and co-workers). The new amide bond in the growing peptide chain is formed by attack of the α-amino group in the new amino acid, brought to the A site of the ribosome by its tRNA, on the acyl carbon linkage of the peptide held at the P site by its tRNA. Acid–base catalysis by groups in the ribosome facilitate the reaction. (Reprinted with permission from Nissen et al., SCIENCE 289:920–930 (2000). Copyright 2000 AAAS. Also reprinted from Monro, R. E., and Marker, K. A., Ribosome-catalysed reaction of puromycin with a formylmethionine containing oligonucleotide, *J. Mol. Biol.* 25 pp. 347–350. Copyright 1967, with permission of Elsevier.)

25.5C Transfer RNA

Transfer RNA has a very low molecular weight when compared to those of mRNA and rRNA. Transfer RNA, consequently, is much more soluble than mRNA or rRNA and is sometimes referred to as soluble RNA. The function of tRNA is to transport amino acids to specific areas on the mRNA bound to the ribosome. There are, therefore, many forms of tRNA, more than one for each of the 20 amino acids that is incorporated into proteins, including the redundancies in the **genetic code** (see Table 25.2).*

The structures of most tRNAs have been determined. They are composed of a relatively small number of nucleotide units (70–90 units) folded into several loops or arms through base

*Although proteins are composed of 22 different amino acids, protein synthesis requires only 20. Proline is converted to hydroxyproline and cysteine is converted to cystine after synthesis of the polypeptide chain has taken place.

TABLE 25.2 The Messenger RNA Genetic Code

Amino Acid	mRNA Base Sequence 5′ → 3′	Amino Acid	mRNA Base Sequence 5′ → 3′	Amino Acid	mRNA Base Sequence 5′ → 3′
Ala	GCA	His	CAC	Ser	AGC
	GCC		CAU		AGU
	GCG	Ile	AUA		UCA
	GCU		AUC		UCG
Arg	AGA		AUU		UCC
	AGG	Leu	CUA		UCU
	CGA		CUC	Thr	ACA
	CGC		CUG		ACC
	CGG		CUU		ACG
	CGU		UUA		ACU
Asn	AAC		UUG	Trp	UGG
	AAU	Lys	AAA	Tyr	UAC
Asp	GAC		AAG		UAU
	GAU	Met	AUG	Val	GUA
Cys	UGC	Phe	UUU		GUG
	UGU		UUC		GUC
Gln	CAA	Pro	CCA		GUU
	CAG		CCC	Chain initiation	
Glu	GAA		CCG	fMet (*N*-formyl-methionine)	AUG
	GAG		CCU		
Gly	GGA			Chain termination	UAA
	GGC				UAG
	GGG				UGA
	GGU				

pairing along the chain (Fig. 25.15). One arm always terminates in the sequence cytosine–cytosine–adenine (CCA). It is to this arm that a specific amino acid becomes attached *through an ester* linkage to the 3′-OH of the terminal adenosine. This attachment reaction is catalyzed by an enzyme that is specific for the tRNA and for the amino acid. The specificity may grow out of the enzyme's ability to recognize base sequences along other arms of the tRNA.

At the loop of still another arm is a specific sequence of bases, called the **anticodon**. The anticodon is highly important because it allows the tRNA to bind with a specific site—called the **codon**—of mRNA. The order in which amino acids are brought by their tRNA units to the mRNA strand is determined by the sequence of codons. This sequence, therefore, constitutes a genetic message. Individual units of that message (the individual words, each corresponding to an amino acid) are triplets of nucleotides.

25.5D The Genetic Code

The triplets of nucleotides (the codons) on mRNA are the genetic code (see Table 25.2). The code must be in the form of three bases, not one or two, because there are 20 different amino acids used in protein synthesis but there are only four different bases in mRNA. If only two bases were used, there would be only 4^2, or 16, possible combinations, a number too small to accommodate all of the possible amino acids. However, with a three-base code, 4^3, or 64, different sequences are possible. This is far more than are needed, and it allows for multiple ways of specifying an amino acid. It also allows for sequences that punctuate protein synthesis, sequences that say, in effect, "start here" and "end here."

Both methionine (Met) and *N*-formylmethionine (fMet) have the same mRNA code (AUG); however, *N*-formylmethionine is carried by a different tRNA from that which car-

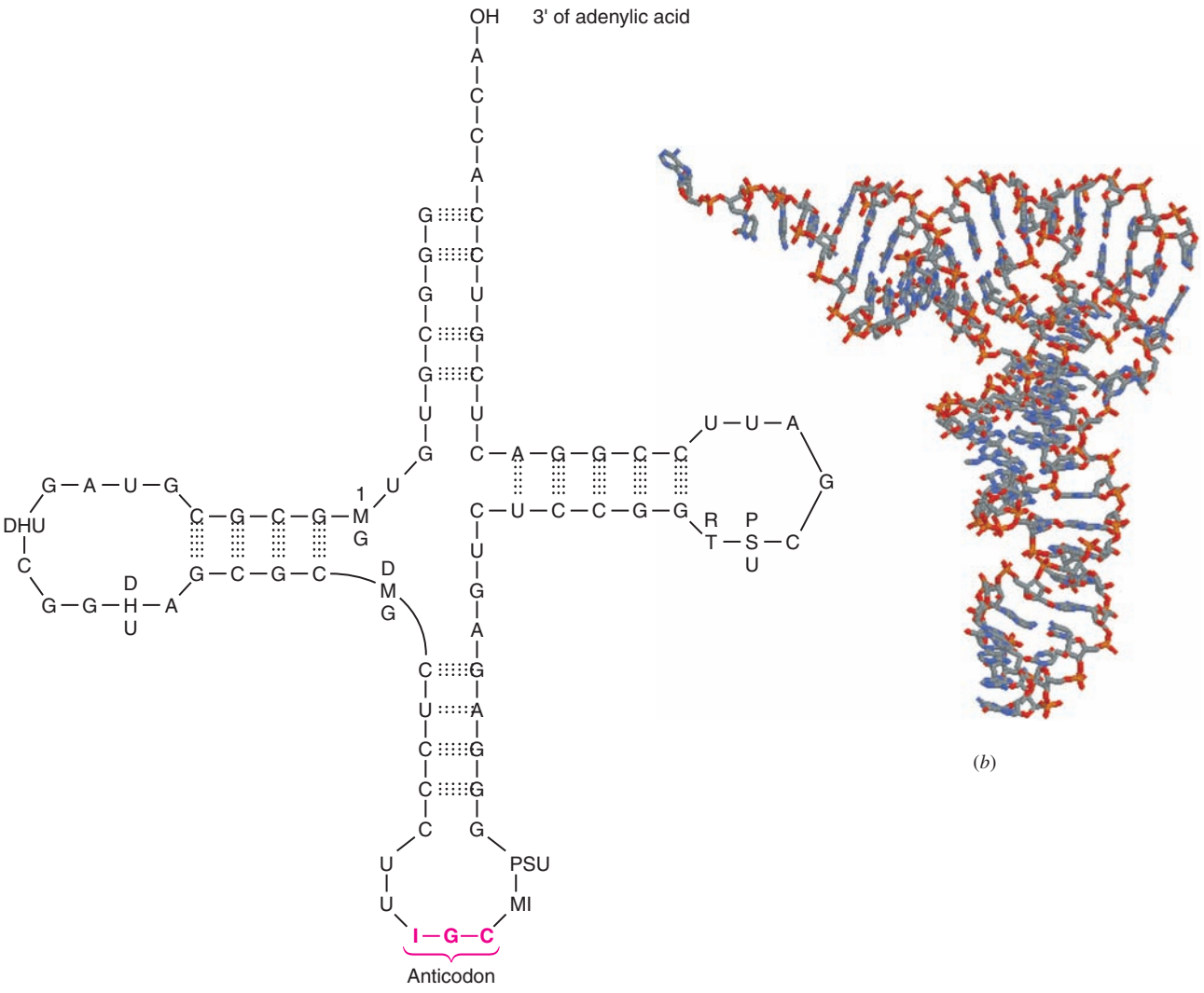
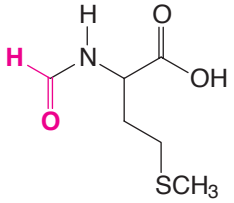


Figure 25.15 *(a)* Structure of a tRNA isolated from yeast that has the specific function of transferring alanine residues. Transfer RNAs often contain unusual nucleosides. PSU = pseudouridine, RT = ribothymidine, MI = 1-methylinosine, I = inosine, DMG = N^2-methylguanosine, DHU = 4,5-dihydrouridine, 1MG = 1-methylguanosine. *(b)* The X-ray crystal structure of a phenylalanine–tRNA from yeast. (For part b, Protein Data Bank PDB ID: 4TNA, http://www.pdb.org. Reprinted from Hingerty, E., Brown, R.S., Jack, A., Further refinement of the structure of yeast $tRNA_{Phe}$, *J. Mol. Biol.* **124**, p. 523. Copyright 1978, with permission of Elsevier.)

ries methionine. *N*-Formylmethionine appears to be the first amino acid incorporated into the chain of proteins in bacteria, and the tRNA that carries fMet appears to be the punctuation mark that says "start here." Before the polypeptide synthesis is complete, *N*-formylmethionine is removed from the protein chain by an enzymatic hydrolysis.



***N*-Formylmethionine (fMet)**

The genetic code can be expressed in mRNA codons (as we have shown in Table 25.2) or in DNA codons. We have chosen to show the mRNA codons because these are the codons that are actually read during the synthesis of polypeptides (the process called **translation** that we discuss next). However, each mRNA molecule (Section 25.5A) acquires its sequence of nucleotides by **transcription** from the corresponding gene of DNA. In transcription, RNA polymerase (along with other transcription factors) opens the DNA double helix and begins the process.

As RNA polymerase transcribes DNA to mRNA, it moves along the complementary strand of DNA reading it in the 3′ to 5′ direction (called the antisense direction), making an mRNA transcript that is the same as the sense strand (the 5′ to 3′ direction) of the DNA (except that uracil replaces thymine). For example:

Sense strand of DNA	5′... CAT	CGT	TTG	ACC	GAT ... 3′
Antisense strand of DNA	3′... GTA	GCA	AAC	TGG	CTA ... 5′
	⇓ Transcription of antisense strand				
mRNA	5′... CAU	CGU	UUG	ACC	GAU ... 3′
	⇓ Translation of mRNA				
Peptide	... His —	Arg —	Leu —	Thr —	Asp ...

Because the synthesis of mRNA proceeds in the 5′ to 3′ direction, the codons for the sense strand of DNA (with the exception of thymine replacing uracil) are the same as those for the mRNA. For example, one DNA codon for valine is GTA. The corresponding mRNA codon for valine is GUA.

25.5E Translation

We are now in a position to see how the synthesis of a hypothetical polypeptide might take place. This process is called **translation**. Let us imagine that a long strand of mRNA is in the cytoplasm of a cell and that it is in contact with ribosomes. Also in the cytoplasm are the 20 different amino acids, each acylated to its own specific tRNA.

As shown in Fig. 25.16, a tRNA bearing fMet uses its anticodon to associate with the proper codon (AUG) on that portion of mRNA that is in contact with a ribosome. The next triplet of bases on the mRNA chain in this figure is AAA; this is the codon that specifies lysine. A lysyl-tRNA with the matching anticodon UUU attaches itself to this site. The two amino acids, fMet and Lys, are now in the proper position for the 50S ribosome subunit to catalyze the formation of an amide bond between them, as shown in Fig. 25.16 (by the mechanism in Fig. 25.14). After this happens, the ribosome moves down the chain so that it is in contact with the next codon. This one, GUA, specifies valine. A tRNA bearing valine (and with the proper anticodon) binds itself to this site. Another peptide bond–forming reaction takes place attaching valine to the polypeptide chain. Then the whole process repeats itself again and again. The ribosome moves along the mRNA chain, other tRNAs move up with their amino acids, new peptide bonds are formed, and the polypeptide chain grows. At some point an enzymatic reaction removes fMet from the beginning of the chain. Finally, when the chain is the proper length, the ribosome reaches a punctuation mark, UAA, saying "stop here." The ribosome separates from the mRNA chain and so, too, does the protein.

Even before the polypeptide chain is fully grown, it begins to form its own specific secondary and tertiary structure. This happens because its primary structure is correct—its amino acids are ordered in just the right way. Hydrogen bonds form, giving rise to

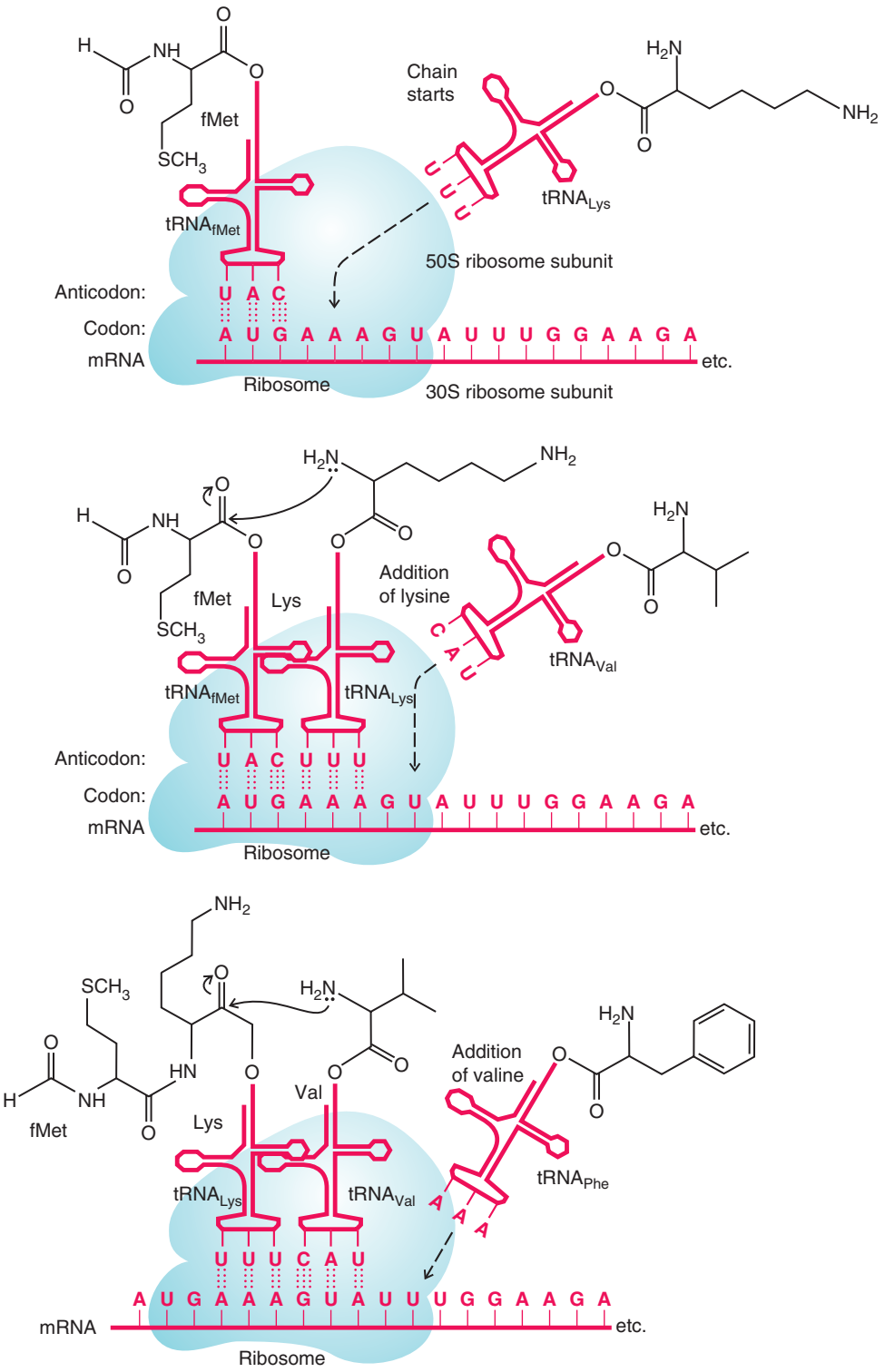


Figure 25.16 Step-by-step growth of a polypeptide chain with mRNA acting as a template. Transfer RNAs carry amino acid residues to the site of mRNA that is in contact with a ribosome. Codon–anticodon pairing occurs between mRNA and RNA at the ribosomal surface. An enzymatic reaction joins the amino acid residues through an amide linkage. After the first amide bond is formed, the ribosome moves to the next codon on mRNA. A new tRNA arrives, pairs, and transfers its amino acid residue to the growing peptide chain, and so on.

specific segments of α helix, pleated sheet, and coil or loop. Then the whole chain folds and bends; enzymes install disulfide linkages, so that when the chain is fully grown, the whole protein has just the shape it needs to do its job. (Predicting 2° and 3° protein structure from amino acid sequence, however, remains a critical problem in structural biochemistry.)

In the meantime, other ribosomes nearer the beginning of the mRNA chain are already moving along, each one synthesizing another molecule of the polypeptide. The time required to synthesize a protein depends, of course, on the number of amino acid residues it contains, but indications are that each ribosome can cause 150 peptide bonds to be formed each minute. Thus, a protein, such as lysozyme, with 129 amino acid residues requires less than a minute for its synthesis. However, if four ribosomes are working their way along a single mRNA chain, a protein molecule can be produced every 13 s.

But why, we might ask, is all this protein synthesis necessary—particularly in a fully grown organism? The answer is that proteins are not permanent; they are not synthesized once and then left intact in the cell for the lifetime of the organism. They are synthesized when and where they are needed. Then they are taken apart, back to amino acids; enzymes disassemble enzymes. Some amino acids are metabolized for energy; others—new ones—come in from the food that is eaten, and the whole process begins again.

Review Problem 25.10

The sense strand of a segment of DNA has the following sequence of bases:

5′ . . . T G G G G G T T T T A C A G C . . . 3′

(a) What mRNA sequence would result from this segment?

(b) Assume that the first base in this mRNA is the beginning of a codon. What order of amino acids would be translated into a polypeptide synthesized along this segment?

(c) Give anticodons for each tRNA associated with the translation in part (b).

Review Problem 25.11

(a) Using the first codon given for each amino acid in Table 25.2, write the base sequence of mRNA that would translate the synthesis of the following pentapeptide:

Arg · Ile · Cys · Tyr · Val

(b) What base sequence in the DNA sense strand would correspond with this mRNA?

(c) What anticodons would appear in the tRNAs involved in the pentapeptide synthesis?

Solved Problem 25.2

Explain how an error of a single base in each strand of DNA could bring about the amino acid error that causes sickle-cell anemia (see "The Chemistry of . . ." box in Section 24.6B).

STRATEGY AND ANSWER A change from GAA to GTA in DNA would lead to a change in mRNA from GAA to GUA (see Table 25.2). This change would result in the glutamic acid residue at position 6 in normal hemoglobin becoming valine (as it is in persons with sickle-cell anemia). Alternatively, a change from GAG to GTG in DNA would lead to a change in mRNA from GAG to GUG that would also result in valine replacing glutamic acid.

25.6 Determining the Base Sequence of DNA: The Chain-Terminating (Dideoxynucleotide) Method

Certain aspects of the strategy used to sequence DNA resemble the methods used to sequence proteins. Both types of molecules require methods amenable to lengthy polymers, but in the case of DNA, a single DNA molecule is so long that it is absolutely necessary to cleave it into smaller, manageable fragments. Another similarity between DNA and proteins is that small sets of molecular building blocks comprise the structures of each, but in the case of DNA, only four nucleotide monomer units are involved instead of the 20 amino acid building blocks used to synthesize proteins. Finally, both proteins and nucleic acids are charged molecules that can be separated on the basis of size and charge using chromatography.

The first part of the process is accomplished by using enzymes called **restriction endonucleases**. These enzymes cleave double-stranded DNA at specific base sequences. Several hundred restriction endonucleases are now known. One, for example, called *Alu*I, cleaves the sequence AGCT between G and C. Another, called *Eco*R1, cleaves GAATTC between G and A. Most of the sites recognized by restriction enzymes have sequences of base pairs with the same order in both strands when read from the 5′ direction to the 3′ direction. For example,

$$5' \leftarrow G—A—A—T—T—C \rightarrow 3'$$
$$3' \leftarrow C—T—T—A—A—G \rightarrow 5'$$

Such sequences are known as **palindromes**. (Palindromes are words or sentences that read the same forward or backward. Examples are "radar" and "Madam, I'm Adam.")

Sequencing of the fragments (often called restriction fragments) can be done chemically or with the aid of enzymes. The first chemical method was introduced by A. Maxam and W. Gilbert (both of Harvard University); the **chain-terminating (dideoxynucleotide) method** was introduced in the same year by F. Sanger (Cambridge University). Essentially all DNA sequencing is currently done using an automated version of the chain-terminating method, which involves enzymatic reactions and 2′,3′-dideoxynucleotides.

Gilbert and Sanger shared the Nobel Prize in Chemistry in 1980 with Paul Berg for their work on nucleic acids. Sanger (Section 24.5B), who pioneered the sequencing of proteins, had won an earlier Nobel prize in 1958 for the determination of the structure of insulin.

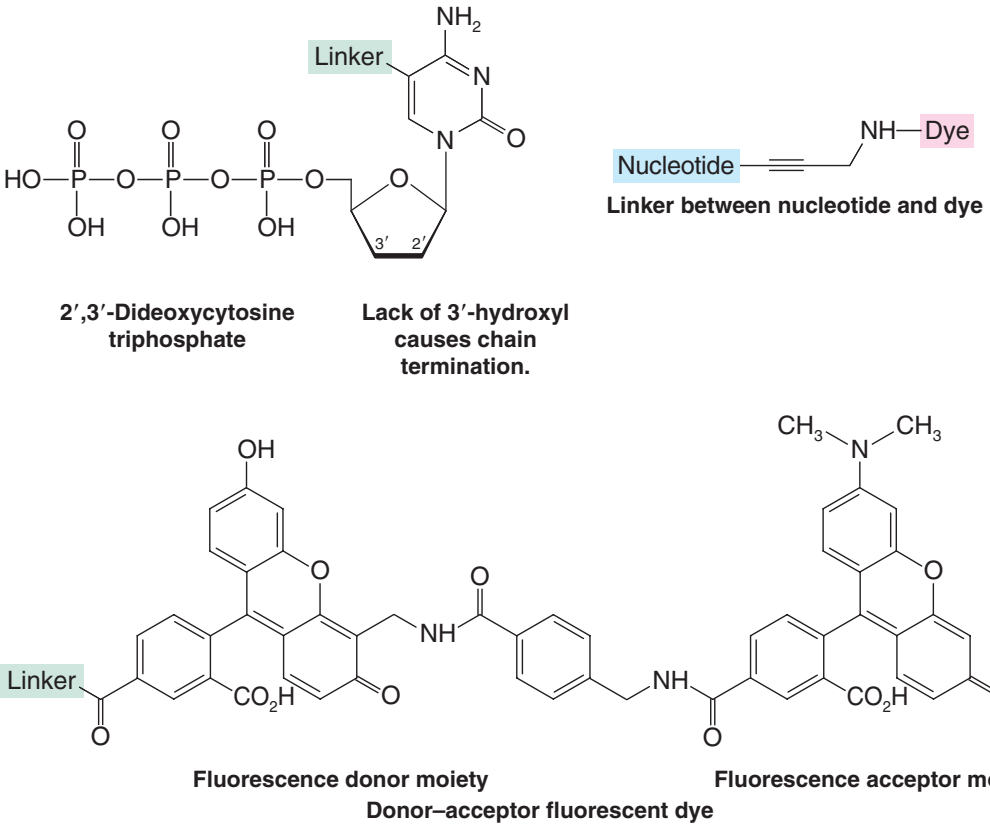
25.6A DNA Sequencing by the Chain-Terminating (Dideoxynucleotide) Method

The chain-terminating method for sequencing DNA involves replicating DNA in a way that generates a family of partial copies that differ in length by one base pair. These partial copies of the parent DNA are separated according to length, and the terminal base on each strand is detected by a covalently attached fluorescent marker.

The mixture of partial copies of the target DNA is made by "poisoning" a replication reaction with a low concentration of unnatural nucleotides. The unnatural terminating nucleotides are 2′,3′-dideoxy analogues of the four natural nucleotides. Lacking the 3′-hydroxyl, each 2′,3′-dideoxynucleotide incorporated is incapable of forming a phosphodiester bond between its 3′ carbon and the next nucleotide that would be needed to continue the polymerization, and hence the chain terminates. Because a low concentration of the dideoxynucleotides is used, only occasionally is a dideoxynucleotide incorporated at random into the growing chains, and thus DNA molecules of essentially all different lengths are synthesized from the parent DNA.

Each terminating dideoxynucleotide is labeled with a fluorescent dye that gives a specific color depending on the base carried by that terminating nucleotide. (An alternate method is to label the *primer*, a short oligonucleotide sequence used to initiate replication of the specific DNA, with specific fluorescent dyes, instead of the dideoxynucleotide terminators, but the general method is the same.) One of the dye systems in use (patented by ABI) consists of a donor chromophore that is initially excited by the laser and which then transfers its energy to an acceptor moiety which produces the observed fluorescence. The donor is tethered to the dideoxynucleotide by a short linker.

A 2′3′-dideoxynucleotide, linker, and fluorescent dye moiety like those used in fluorescence-tagged dideoxynucleotide DNA sequencing reactions



The replication reaction used to generate the partial DNA copies is similar but not identical to the polymerase chain reaction (PCR) method (Section 25.8). In the dideoxy sequencing method only one primer sequence of DNA is used, and hence only one strand of the DNA is copied, whereas in the PCR, two primers are used and both strands are copied simultaneously. Furthermore, in sequencing reactions the chains are deliberately terminated by addition of the dideoxy nucleotides.

Capillary electrophoresis is the method most commonly used to separate the mixture of partial DNAs that results from a sequencing reaction. Capillary electrophoresis separates the DNAs on the basis of size and charge, allowing nucleotides that differ by only one base length to be resolved. Computerized acquisition of fluorescence data as the differently terminated DNAs pass the detector generates a four-color chromatogram, wherein each consecutive peak represents a DNA molecule one nucleotide longer than the previous one. The color of each peak represents the terminating nucleotide in that molecule. Since each of the four types of dideoxy terminating bases fluoresces a different color, the sequence of nucleotides in the DNA can be read directly. An example of sequence data from this kind of system is shown in Fig. 25.17.

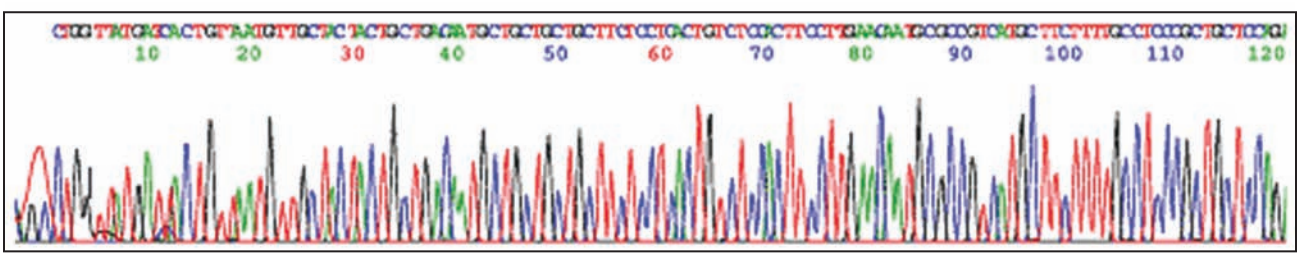


Figure 25.17 Example of data from an automated DNA sequencer. (*Courtesy of Applied Biosystems.*)

Use of automated methods for DNA sequencing represents an exponential increase in speed over manual methods employing vertical slab polyacrylamide gel electrophoresis (Fig. 25.18). Only a few thousand bases per day (at most) could be sequenced by a person using the manual method. Now it is possible for a single machine running parallel and continuous analyses to sequence almost 3 million bases per day using automated capillary electrophoresis and laser fluorescence detection. As an added benefit, the ease of DNA sequencing often makes it easier to determine the sequence of a protein by the sequence of all or part of its corresponding gene, rather than by sequencing the protein itself (see Section 24.5).

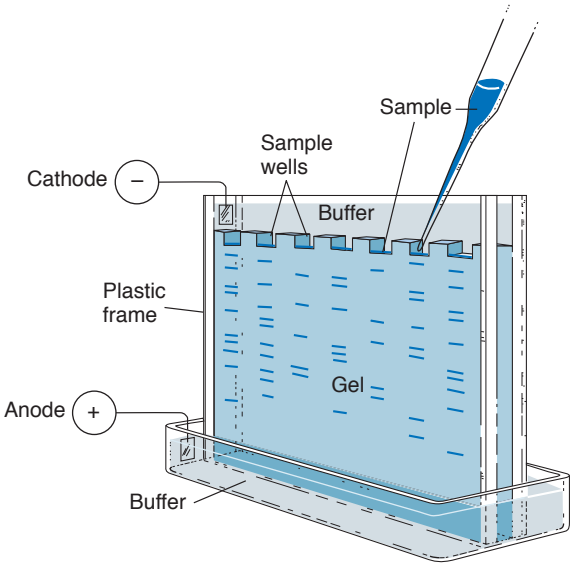


Figure 25.18 An apparatus for gel electrophoresis. Samples are applied in the slots at the top of the gel. Application of a voltage difference causes the samples to move. The samples move in parallel lanes. (Reprinted with permission of John Wiley & Sons, Inc., from Voet, D. and Voet, J. G., *Biochemistry*, Second Edition. © 1995 Voet, D. and Voet, J. G.)

The development of high-throughput methods for sequencing DNA is largely responsible for the remarkable success achieved in the Human Genome Project. Sequencing the 3 billion base pairs in the human genome could never have been completed before 2003 and the 50th anniversary of Watson and Crick's elucidation of the structure of DNA had high-throughput sequencing methods not come into existence.*

25.7 Laboratory Synthesis of Oligonucleotides

Synthetic oligonucleotides are needed for a variety of purposes. One of the most important and common uses of synthetic oligonucleotides is as primers for nucleic acid sequencing and for PCR (Section 25.8). Another important application is in the research and development of **antisense oligonucleotides**, which hold potential as therapies for a variety of diseases. An antisense oligonucleotide is one that has a sequence complementary to the coding sequence in a DNA or RNA molecule. Synthetic oligonucleotides that bind tightly to DNA or mRNA sequences from a virus, bacterium, or other disease condition may be able to shut down expression of the target protein associated with those conditions. For example, if the sense portion of DNA in a gene reads

A—G—A—C—C—G—T—G—G

the antisense oligonucleotide would read

T—C—T—G—G—C—A—C—C

*The Human Genome Project website of the U.S. Department of Energy provides a wealth of resources for further information: www.ornl.gov/hgmis/.

The ability to deactivate specific genes in this way holds great medical promise. Many viruses and bacteria, during their life cycles, use a method like this to regulate some of their own genes. The hope, therefore, is to synthesize antisense oligonucleotides that will seek out and destroy viruses in a person's cells by binding with crucial sequences of the viral DNA or RNA. Synthesis of such oligonucleotides is an active area of research today and is directed at many viral diseases, including AIDS, as well as lung and other forms of cancer.

Current methods for **oligonucleotide synthesis** are similar to those used to synthesize proteins, including the use of automated solid-phase techniques (Section 24.7D). A suitably protected nucleotide is attached to a solid phase called a "controlled pore glass," or CPG (Fig. 25.19), through a linkage that can ultimately be cleaved. The next protected nucleotide in the form of a **phosphoramidite** is added, and coupling is brought about by a coupling agent, usually 1,2,3,4-tetrazole. The phosphite triester that results from the coupling is oxidized to phosphate triester with iodine, producing a chain that has been lengthened by one nucleotide. The **dimethoxytrityl (DMTr)** group used to protect the 5′ end of the added nucleotide is removed by treatment with acid, and the steps **coupling, oxidation, detritylation**, as shown in Figure 25.19, are repeated. (All the steps are carried out in nonaqueous solvents.) With automatic synthesizers the process can be repeated at least 50 times and the time for a complete cycle is 40 min or less. The synthesis is monitored by spectrophotometric detection of the dimethoxytrityl cation as it is released in each cycle (much like the monitoring of Fmoc release in solid-phase peptide synthesis). After the desired oligonucleotide has been synthesized, it is released from the solid support and the various protecting groups, including those on the bases, are removed.

25.8 *The Polymerase Chain Reaction*

The **polymerase chain reaction (PCR)** is an extraordinarily simple and effective method for exponentially multiplying (amplifying) the number of copies of a DNA molecule. Beginning with even just a single molecule of DNA, the PCR can generate 100 billion copies in a single afternoon. The reaction is easy to carry out: It requires only a miniscule sample of the target DNA (picogram quantities are sufficient), a supply of nucleotide triphosphate reagents and primers to build the new DNA, DNA polymerase to catalyze the reaction, and a device called a thermal cycler to control the reaction temperature and automatically repeat the reaction. The PCR has had a major effect on molecular biology. Perhaps its most important role has been in the sequencing of the human genome (Sections 25.6 and 25.9), but now virtually every aspect of research involving DNA involves the PCR at some point.

One of the original aims in developing the PCR was to use it in increasing the speed and effectiveness of prenatal diagnosis of sickle-cell anemia (Section 24.6B). It is now being applied to the prenatal diagnosis of a number of other genetic diseases, including muscular dystrophy and cystic fibrosis. Among infectious diseases, the PCR has been used to detect cytomegalovirus and the viruses that cause AIDS, certain cervical carcinomas, hepatitis, measles, and Epstein–Barr disease.

The PCR is a mainstay in forensic sciences as well, where it may be used to copy DNA from a trace sample of blood or semen or a hair left at the scene of a crime. It is also used in evolutionary biology and anthropology, where the DNA of interest may come from a 40,000-year-old woolly mammoth or the tissue of a mummy. It is also used to match families with lost relatives (see the chapter opening vignette). There is almost no area with biological significance that does not in some way have application for use of the PCR reaction.

Mullis was awarded the Nobel Prize in Chemistry for this work in 1993.

The PCR was invented by Kary B. Mullis and developed by him and his co-workers at Cetus Corporation. It makes use of the enzyme DNA polymerase, discovered in 1955 by Arthur Kornberg and associates at Stanford University. In living cells, DNA polymerases

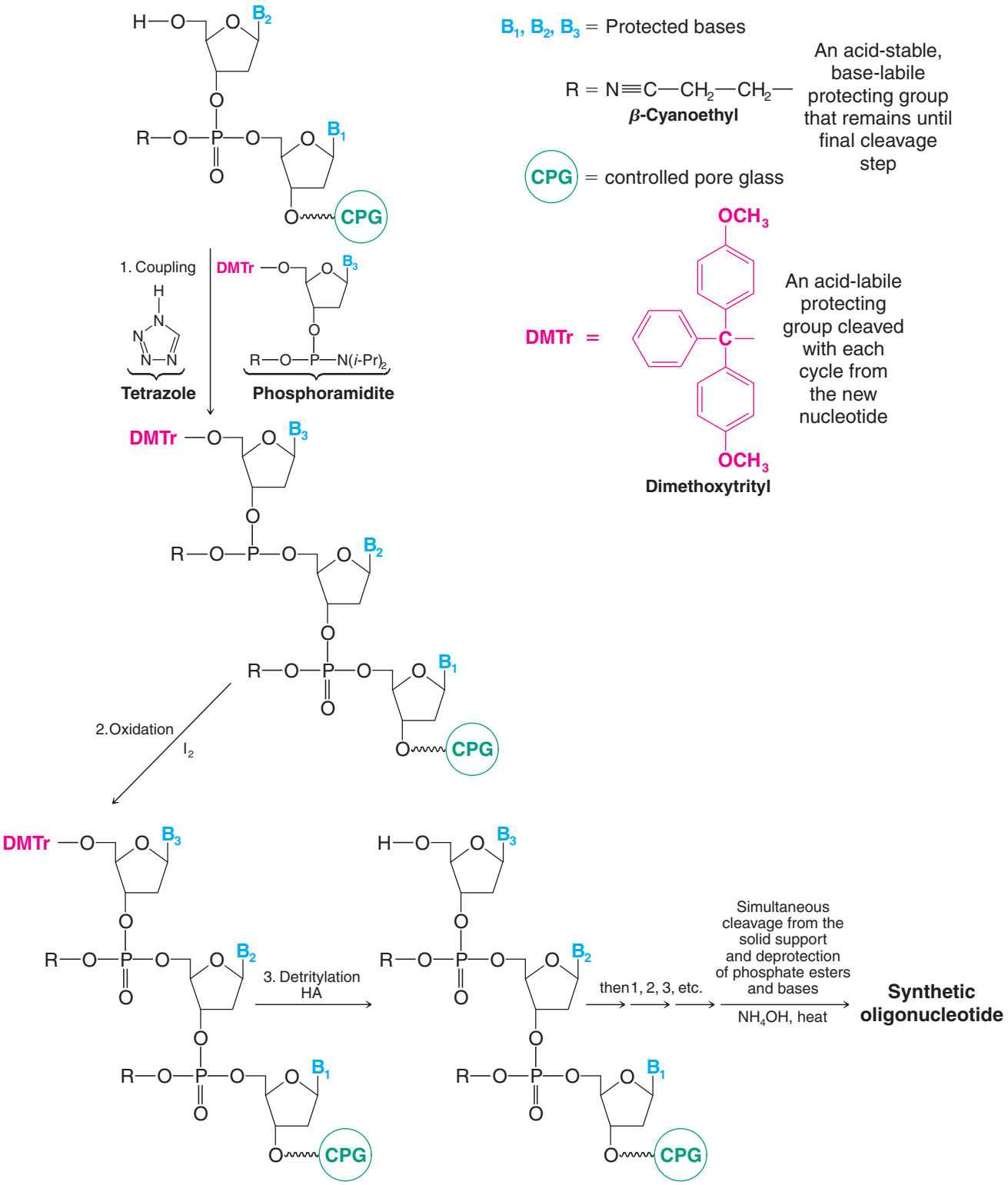


Figure 25.19 The steps involved in automated synthesis of oligonucleotides using the phosphoramidite coupling method.

help repair and replicate DNA. The PCR makes use of a particular property of DNA polymerases: their ability to attach additional nucleotides to a short oligonucleotide "primer" when the primer is bound to a complementary strand of DNA called a template. The nucleotides are attached at the 3′ end of the primer, and the nucleotide that the polymerase attaches will be the one that is complementary to the base in the adjacent position on the template strand. If the adjacent template nucleotide is G, the polymerase adds C to the primer; if the adjacent template nucleotide is A, then the polymerase adds T, and so on. Polymerase repeats this process again and again as long as the requisite nucleotides (as triphosphates) are present in the solution, until it reaches the 5′ end of the template.

Figure 25.20 shows one PCR cycle. The target DNA, a supply of nucleotide triphosphate monomers, DNA polymerase, and the appropriate oligonucleotide primers (one primer sequence for each 5′ to 3′ direction of the target double-stranded DNA) are added to a small reaction vessel. The mixture is briefly heated to approximately 90°C to separate the DNA strands (denaturation); it is cooled to 50–60°C to allow the primer sequences and DNA polymerase to bind to each of the separated strands (annealing); and it is warmed to about 70°C to extend each strand by polymerase-catalyzed condensation of nucleotide triphosphate monomers complementary to the parent DNA strand. Another cycle of the PCR begins by heating to separate the new collection of DNA molecules into single strands, cooling for the annealing step, and so on.

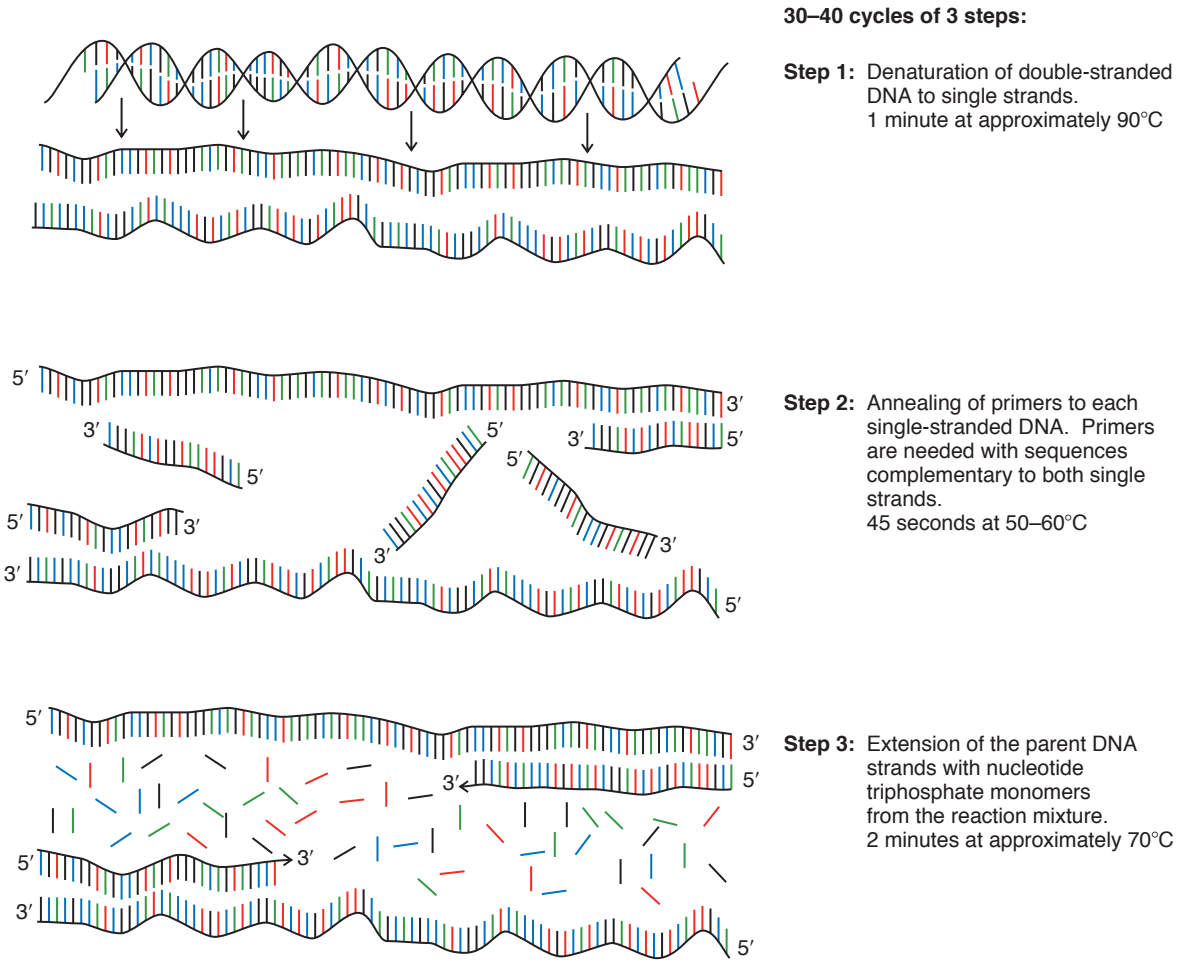


Figure 25.20 One cycle of the PCR. Heating separates the strands of DNA of the target to give two single-stranded templates. Primers, designated to complement the nucleotide sequences flanking the targets, anneal to each strand. DNA polymerase, in the presence of nucleotide triphosphates, catalyzes the synthesis of two pieces of DNA, each identical to the original target DNA. (*Used with permisson from Andy Vierstraete, University of Ghent.*)

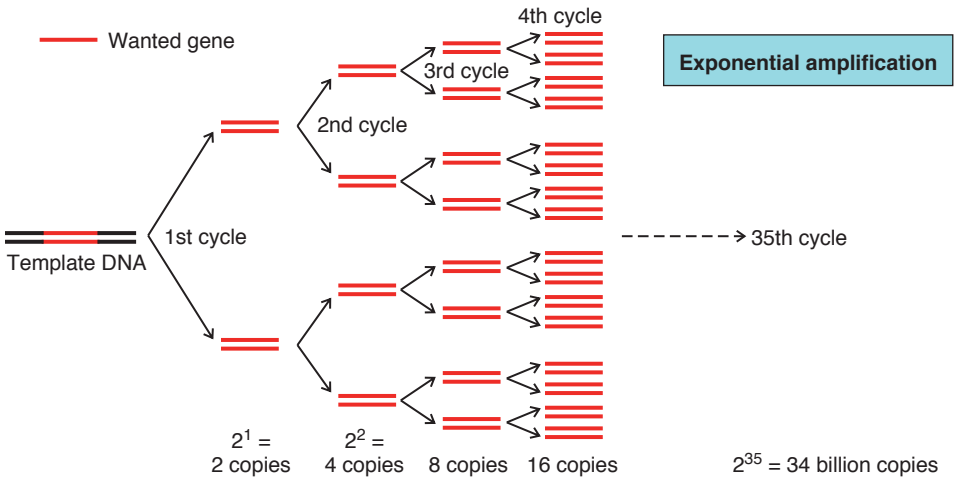


Figure 25.21 Each cycle of the PCR doubles the number of copies of the target area. (*Used with permisson from Andy Vierstraete, University of Ghent.*)

Each cycle, taking only a few minutes, doubles the amount of target DNA that existed prior to that step (Fig. 25.21). The result is an exponential increase in the amount of DNA over time. After n cycles, the DNA will have been replicated 2^n times—after 10 cycles there is roughly 1000 times as much DNA; after 20 cycles roughly 1 million times as much; and so on. Thermal cycling machines can carry out approximately 20 PCR cycles per hour, resulting in billions of DNA copies over a single afternoon.

Each application of PCR requires primers that are 10–20 nucleotides in length and whose sequences are complementary to short, conveniently located sequences flanking the target DNA sequence. The primer sequence is also chosen so that it is near sites that are cleavable with restriction enzymes. Once a researcher determines what primer sequence is needed, the primers are usually purchased from commercial suppliers who synthesize them on request using solid-phase oligonucleotide synthesis methods like that described in Section 25.7.

As an intriguing adjunct to the PCR story, it turns out that cross-fertilization between disparate research fields greatly assisted development of current PCR methods. In particular, the discovery of extremozymes, which are enzymes from organisms that live in high-temperature environments, has been of great use. DNA polymerases now typically used in PCR are heat-stable forms derived from thermophilic bacteria. Polymerases such as Taq polymerase, from the bacterium *Thermus aquaticus*, found in places such as geyser hot springs, and Vent_R™, from bacteria living near deep-sea thermal vents, are used. Use of extremozyme polymerases facilitates PCR by allowing elevated temperatures to be used for the DNA melting step without having to worry about denaturing the polymerase enzyme at the same time. All materials can therefore be present in the reaction mixture throughout the entire process. Furthermore, use of a higher temperature during the chain extension also leads to faster reaction rates. (See "The Chemistry of . . . Stereoselective Reductions of Carbonyl Groups," Section 12.3C, for another example of the use of high-temperature enzymes.)

Thermophilic bacteria, growing in hot springs like these at Yellowstone National Park, produce heat-stable enzymes called extremozymes that have proved useful for a variety of chemical processes.

25.9 *Sequencing of the Human Genome: An Instruction Book for the Molecules of Life*

The announcement by scientists from the public Human Genome Project and Celera Genomics Company in June 2000 that sequencing of the approximately 3 billion base pairs in the human genome was complete marked achievement of one of the most important and ambitious scientific endeavors ever undertaken. To accomplish this feat, data were pooled from thousands of scientists working around the world using tools including PCR (Section 25.8), dideoxynucleotide sequencing reactions (Section 25.6), capillary electrophoresis, laser-induced fluorescence, and supercomputers. What was ultimately produced is a transcript of our chromosomes that could be called an instruction book for the molecules of life.

"This structure has novel features, which are of considerable biological importance." James Watson, one of the scientists who determined the structure of DNA.

But what do the instructions in the genome say? How can we best make use of the molecular instructions for life? Of the roughly 35,000 genes in our DNA, the function of only a small percentage of genes is understood. Discovering genes that can be used to benefit our human condition and the chemical means to turn them on or off presents some of the greatest opportunities and challenges for scientists of today and the future. Sequencing the genome was only the beginning of the story.

As the story unfolds, chemists will continue to add to the molecular archive of compounds used to probe our DNA. DNA microchips, with 10,000 or more short diagnostic sequences of DNA chemically bonded to their surface in predefined arrays, will be used to test DNA samples for thousands of possible genetic conditions in a single assay. With the map of our genome in hand, great libraries of potential drugs will be tested against genetic targets to discover more molecules that either promote or inhibit expression of key gene products. Sequencing of the genome will also accelerate development of molecules that interact with proteins, the products of gene expression. Knowledge of the genome sequence will expedite identification of the genes coding for interesting proteins, thus allowing these proteins to be expressed in virtually limitless quantities. With an ample supply of target proteins available, the challenges of solving three-dimensional protein structures and understanding their functions will also be overcome more easily. Optimization of the structures of small organic molecules that interact with proteins will also occur more rapidly because the protein targets for these molecules will be available faster and in greater quantity. There is no doubt that the pace of research to develop new and useful organic molecules for interaction with gene and protein targets will increase dramatically now that the genome has been sequenced. The potential to use our chemical creativity in the fields of genomics and proteomics is immense.

Key Terms and Concepts

The key terms and concepts that are highlighted in **bold, blue text** within the chapter are defined in the glossary (at the back of the book) and have hyperlinked definitions in the accompanying ***WileyPLUS*** course (www.wileyplus.com).

Problems

Note to Instructors: Many of the homework problems are available for assignment via ***WileyPLUS,*** an online teaching and learning solution.

NUCLEIC ACID STRUCTURE

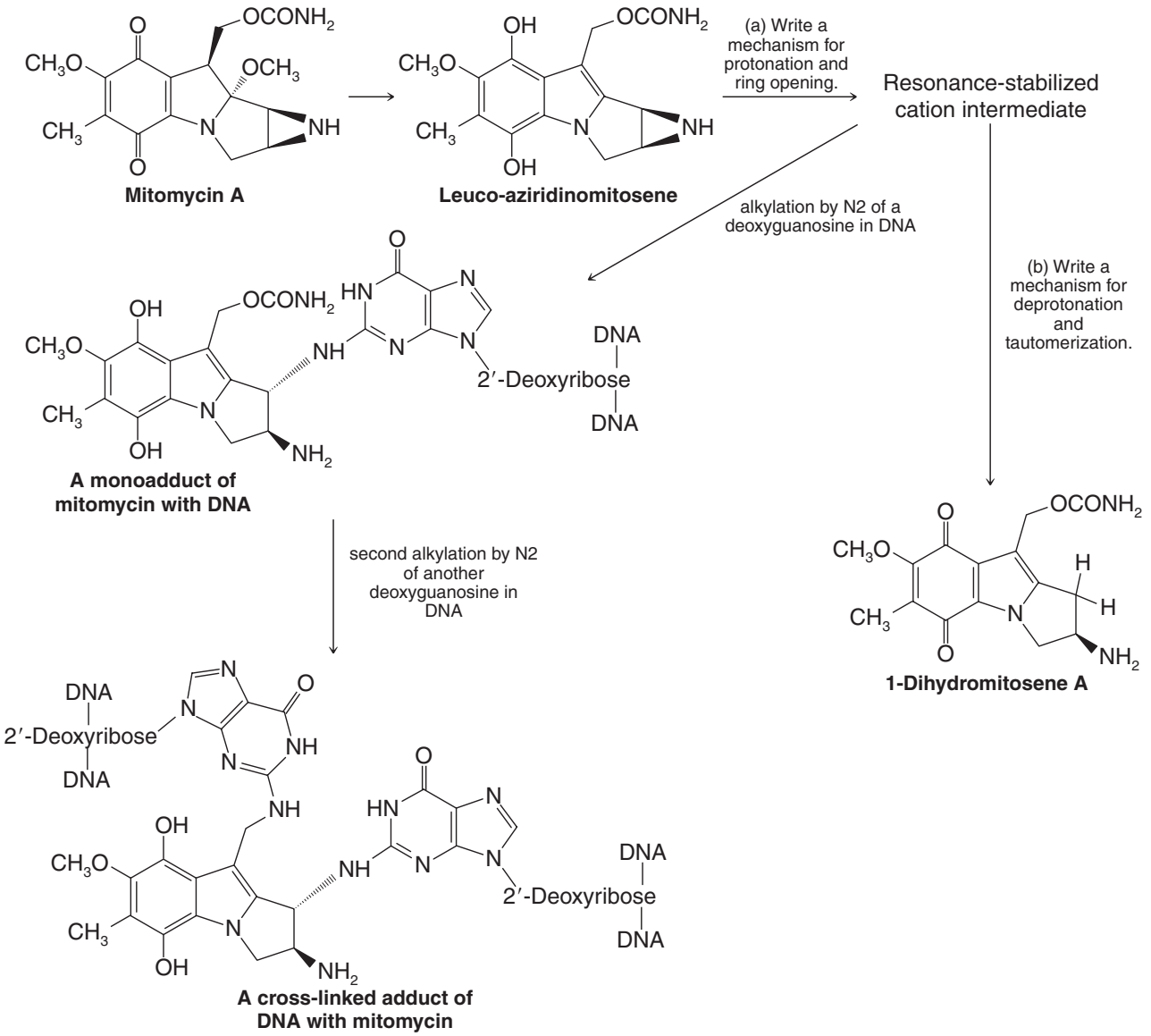
25.12 Write the structure of the RNA dinucleotide G–C in which G has a free 5′-hydroxyl group and C has a free 3′-hydroxyl group.

25.13 Write the structure of the DNA dinucleotide T–A in which T has a free 5′-hydroxyl group and A has a free 3′-hydroxyl group.

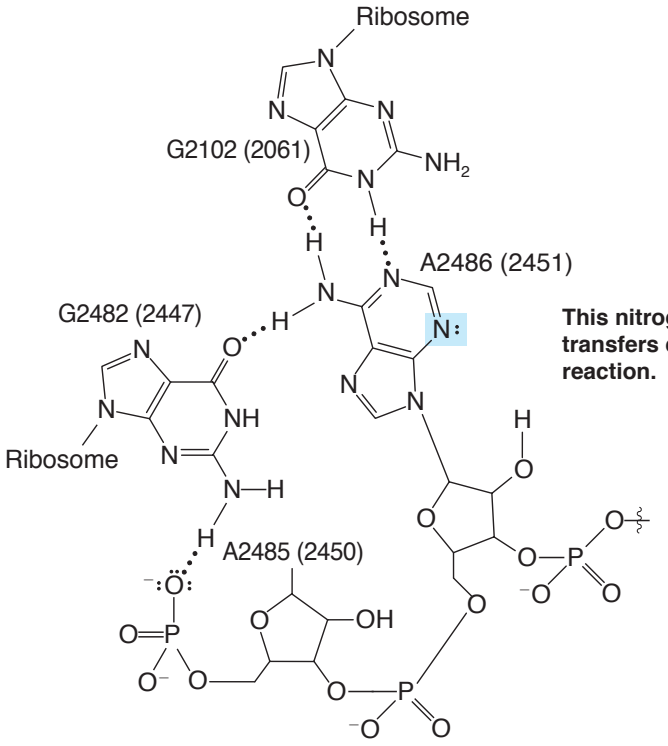
MECHANISMS

25.14 The example of a silyl–Hilbert–Johnson nucleosidation reaction in Section 25.3 is presumed to involve an intermediate ribosyl cation that is stabilized by intramolecular interactions involving the C2 benzoyl group. This intermediate blocks attack by the heterocyclic base from the α face of the ribose ring but allows attack on the β face, as required for formation of the desired product. Propose a structure for the ribosyl cation intermediate that explains the stereoselective bonding of the base.

25.15 **(a)** Mitomycin is a clinically used antitumor antibiotic that acts by disrupting DNA synthesis through covalent bond-forming reactions with deoxyguanosine in DNA. Maria Tomasz (Hunter College) and others have shown that alkylation of DNA by mitomycin occurs by a complex series of mechanistic steps. The process begins with reduction of the quinone ring in mitomycin to its hydroquinone form, followed by elimination of methanol from the adjacent ring to form an intermediate called leuco-aziridinomitosene. One of the paths by which leuco-aziridinomitosene alkylates DNA involves protonation and opening of the three-membered aziridine ring, resulting in an intermediate cation that is resonance stabilized by the hydroquinone group. Attack on the cation by N2 of a deoxyguanosine residue leads to a monoalkylated DNA product, as shown in the scheme. Write a detailed mechanism to show how the ring opening might occur, including resonance forms for the cation intermediate, followed by nucleophilic attack by DNA. (Intra- or interstrand cross-linking of DNA can further occur by reaction of another deoxyguanosine residue to displace the carbamoyl group of the initial mitosene–DNA monoadduct. A cross-linked adduct is also shown.) **(b)** 1-Dihydromitosene A is sometimes formed from the cation intermediate in part (a) by loss of a proton and tautomerization. Propose a detailed mechanism for the formation of 1-dihydromitosene A from the resonance-stabilized cation of part (a).



25.16 As described in Section 25.5B, acid–base catalysis is believed to be the mechanism by which ribosomes catalyze the formation of peptide bonds in the process of protein translation. Key to this proposal is assistance by the N3 nitrogen (highlighted in the scheme below) of a nearby adenine in the ribosome for the removal of a proton from the α-amino group of the amino acid adding to the growing peptide chain (Fig. 25.14). The ability of this adenine group to remove the proton is, in turn, apparently facilitated by relay of charge made possible by other nearby groups in the ribosome. The constellation of these groups is shown in the scheme. Draw mechanism arrows to show formation of a resonance contributor wherein the adenine group could carry a formal negative charge, thereby facilitating its removal of the α-amino proton of the amino acid. (The true electronic structure of these groups is not accurately represented by any single resonance contributor, of course. A hybrid of the contributing resonance structures weighted according to stability would best reflect the true structure.)



Learning Group Problem

Research suggests that expression of certain genes is controlled by conversion of some cytosine bases in the genome to 5-methylcytosine by an enzyme called DNA methyltransferase. Cytosine methylation may be a means by which some genes are turned off as cells differentiate during growth and development. It may also play a role in some cancer processes and in defending the genome from foreign DNA such as viral genes. Measuring the level of methylation in DNA is an important analytical process. One method for measuring cytosine methylation is known as methylation-specific PCR. This technique requires that all unmethylated cytosines in a sample of DNA be converted to uracil by deamination of the C4 amino group in the unmethylated cytosines. This is accomplished by treating the DNA with sodium bisulfite ($NaHSO_3$) to form a bisulfite addition product with its unmethylated cytosine residues. The cytosine sulfonates that result are then subjected to hydrolysis conditions that convert the C4 amino group to a carbonyl group, resulting in uracil sulfonate. Finally, treatment with base causes elimination of the sulfonate group to produce uracil. The modified DNA is then amplified by PCR using primers designed to distinguish DNA with methylated cytosine from cytosine-to-uracil converted bases.

Write detailed mechanisms for the reactions used to convert cytosine to uracil by the above sequence of steps.

Chapter 1

1.18 (a), (c), (f), (g) are tetrahedral; (e) is trigonal planar; (b) is linear; (d) is angular; (h) is trigonal pyramidal.

1.23 (a) and (d); (b) and (e); and (c) and (f).

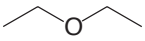
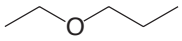
1.31 (a), (g), (i), (l), represent different compounds that are not isomeric; (c–e), (h), (j), (m), (n), (o) represent the same compound; (b), (f), (k), (p) represent constitutional isomers.

1.38 (a) The structures differ in the positions of the nuclei.

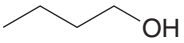
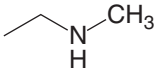
1.46 (a) A negative charge; (b) a negative charge; (c) trigonal pyramidal.

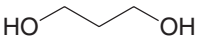
Chapter 2

2.11 (c) Propyl bromide; (d) isopropyl fluoride; (e) phenyl iodide.

2.14 (a)  (b) 

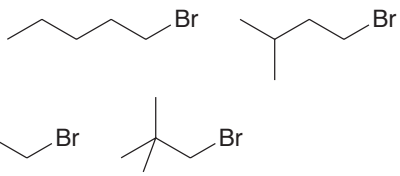
(e) diisopropyl ether.

2.25 (a)  (b) 

(c) 

2.29 (a) ketone; (c) 2° alcohol; (e) 2° alcohol.

2.30 (a) 3 alkene, and a 2° alcohol; (c) phenyl and 1° amine; (e) phenyl, ester and 3° amine; (g) alkene and 2 ester groups.

2.35 (f) 

2.54 Ester

Chapter 3

3.2 (a), (c), (d), and (f) are Lewis bases; (b) and (e) are Lewis acids.

3.4 (a) $[H_3O^+] = [HCO_2^-] = .0042\ M$; (b) Ionization = 4.2%.

3.5 (a) $pK_a = 7$; (b) $pK_a = -0.7$; (c) Because the acid with a $pK_a = 5$ has a larger K_a, it is the stronger acid.

3.8 The pK_a of the methylaminium ion is equal to 10.6 (Section 3.6B). Because the pK_a of the anilinium ion is equal to 4.6, the anilinium ion is a stronger acid than the methylaminium ion, and aniline ($C_6H_5NH_2$) is a weaker base than methylamine (CH_3NH_2).

3.14 (a) $CHCl_2CO_2H$ would be the stronger acid because the electron-withdrawing inductive effect of two chlorine atoms would make its hydroxyl proton more positive. (c) CH_2FCO_2H would be the stronger acid because a fluorine atom is more electronegative than a bromine atom and would be more electron withdrawing.

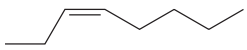
3.28 (a) $pK_a = 3.752$; (b) $K_a = 10^{-13}$.

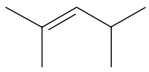
Chapter 4

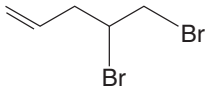
4.8 (a) (1,1-dimethylethyl)cyclopentane or *tert*-butyl-cyclopentane; (c) butylcyclohexane; (e) 2-chlorocyclopentanol.

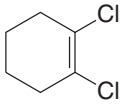
4.9 (a) 2-Chlorobicyclo[1.1.0]butane; (c) bicyclo[2.1.1]hexane; (e) 2-methylbicyclo[2.2.2]octane.

4.10 (a) *trans*-3-Heptene; (c) 4-ethyl-2-methyl-1-hexene

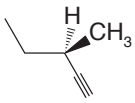
4.11 (a) 

(c) 

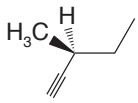
(e) 

(g)  (i) 

4.12 1-Hexyne, 2-Hexyne, 3-Hexyne, 4-Methyl-1-pentyne, 4-Methyl-2-pentyne 3,3-Dimethyl-1-butyne



(*R*)-3-Methyl-1-pentyne

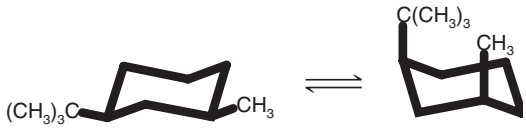


(*S*)-3-Methyl-1-pentyne

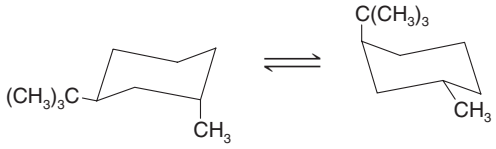
4.24 (a) 3,3,4-Trimethylhexane; (c) 3,5,7-Trimethylnonane; (e) 2-Bromobicyclo[3.3.1]nonane; (g) Cyclobutylcyclopentane.

4.39 (a) Pentane would boil higher because its chain is unbranched. (c) 2-Chloropropane because it is more polar and has a higher molecular weight. (e) CH_3COCH_3 because its molecules are more polar.

4.43

(a) 

More stable conformation because both alkyl groups are equatorial

(b) 

More stable because larger group is equatorial