

WHAT'S AHEAD

- 31.1 ► Amines and the Amide Bond
- 31.2 ► Amino Acids
- 31.3 ► Proteins, Peptides, and Enzymes
- 31.4 ► Nucleic Acids and DNA

31

NITROGEN-CONTAINING ORGANIC COMPOUNDS

31.1 | Amines and the Amide Bond



The discovery of penicillin in 1928 by Alexander Fleming was serendipitous. He had left an uncovered petri dish containing staphylococci bacteria on the windowsill of his laboratory at the St. Mary's Medical School of London University. Mold spores blew in through the open window and started to grow on the culture medium. Some days later, Fleming observed that the bacteria surrounding the mold was dying. He surmised that it was not the mold itself (from the *Penicillium* genus), but a metabolite of the mold that was causing the effect and named it penicillin. Fleming did little to capitalize on his discovery, but Ernst Chain, a biochemist working at Oxford University's Sir William Dunn School of Pathology, as well as Howard Florey, who headed the school, recognized the enormous potential of the discovery to medicine. Work commenced on isolating this "mold juice" and involved fermenting thousands of liters of mold broth in vessels exposed to air. By 1945, Dorothy Hodgkin had determined the chemical structure of penicillin using X-ray crystallography and found it was a relatively small molecule centered around a four membered ring containing a cyclic amide called a β -lactam. The 1945 Nobel prize in physiology or medicine was awarded jointly to Fleming, Florey, and Chain.

Chapters 27–29 introduced the chemistry of oxygen-containing compounds; in this chapter we will examine the complementary role played by nitrogen-containing molecules. In fact, nitrogen-containing organic compounds form an incredibly large and important class in their own right.

By the end of this section, you should be able to

- Name simple amines and amides
- Recognize several methods by which amines may be synthesized
- Understand the acid-base properties of amines

Amines

Amines are organic bases with the general formula



where R, R' and R'' may be H or a hydrocarbon group. Amines containing alkyl groups are called **aliphatic amines**. Aniline is the parent molecule of a class of amines known as **aromatic amines**. This class is characterized by a covalent bond between the nitrogen atom and at least one aromatic group (for example, phenyl or naphthyl). Amines that form part of a cyclic structure are called **cyclic amines** or *heterocycles*. A special class of heterocyclic compounds contains a nitrogen atom within an aromatic ring (for example, pyridine). We also class these amines as aromatic amines. In many cases the lone pair found on nitrogen in an aromatic amine can contribute to the aromaticity. **Figure 31.1** gives some examples of amine classes, including the five- and six-membered heterocycles pyrrolidine and piperidine, and the aromatic heterocycles pyrrole and pyridine.

Amines are further classified as primary, secondary or tertiary.

- **Primary (1°) amine:** contains *one* alkyl or aryl group bonded to nitrogen.
- **Secondary (2°) amine:** contains *two* alkyl or aryl groups bonded to nitrogen.
- **Tertiary (3°) amine:** contains *three* alkyl or aryl groups bonded to nitrogen.
- **Quaternary (4°) ammonium salts:** contain *four* alkyl or aryl groups directly attached to nitrogen.

Go Figure			
What makes an amine secondary?			
$\text{CH}_3\text{CH}_2\ddot{\text{N}}\text{H}_2$	$(\text{CH}_3\text{CH}_2)_2\ddot{\text{N}}\text{H}$	$(\text{CH}_3\text{CH}_2)_3\ddot{\text{N}}$	$(\text{CH}_3\text{CH}_2)_4\text{N}^+ \text{I}^-$
ethylamine aliphatic amine	diethylamine aliphatic amine	triethylamine aliphatic amine	tetraethylammonium iodide ammonium salt
1° amine	2° amine	3° amine	4° ammonium salt
piperidine aliphatic amine	pyrrolidine aliphatic amine	pyridine aromatic amine	pyrrole aromatic amine
heterocyclic 2° amine	heterocyclic 2° amine	heterocyclic 3° amine	heterocyclic 2° amine

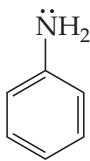
▲ **Figure 31.1** Amine classes.

Figure 31.2 shows the structure of some industrially and biologically important amines. Aniline, for example, is an aromatic amine used extensively in the manufacture of dye stuffs. A significant number of amines are derived from plant material, such as roots, leaves and bark. Amines that are derived from plant material and contain a basic nitrogen atom are commonly called **alkaloids**. Alkaloids form an enormous class of compounds whose physiological properties have been studied intensively for over 150 years. Coniine, for example, is a highly toxic alkaloid, found to be the active ingredient in “poison hemlock”, which was responsible for the death of Socrates. Nicotine is a highly addictive alkaloid found in tobacco leaves. Codeine and morphine, are both analgesics and alkaloids extracted from opium poppies. Adrenaline is a principal compound in a group of compounds known as the *phenylethylamines*. Adrenaline is not an alkaloid; it is a *hormone* that is released into the body in response to stress. It causes increased blood flow to the muscles and brain, accelerates respiration and heart rate, and stimulates the release of stored energy into the blood. When injury to the body occurs, it potentiates the healing process. At times, adrenaline also functions as a neurotransmitter. Pyridoxamine, or vitamin B₆, is a water soluble vitamin that exists in three major chemical forms: pyridoxine, pyridoxal

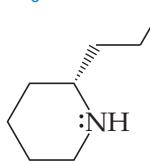


Go Figure

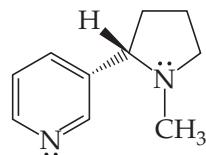
Can you identify which amine is 1° and which is 3° in vitamin B₆?



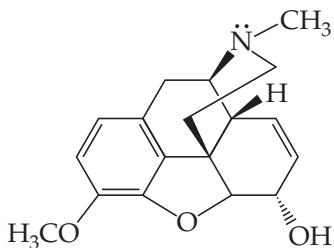
aniline
(1° amine)
aromatic amine



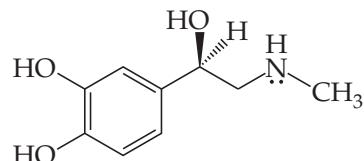
coniine
(2° amine)
heterocyclic amine



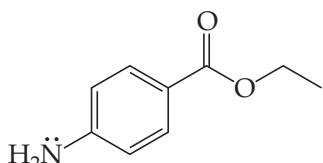
nicotine
(3° amine)
heterocyclic and aromatic amine



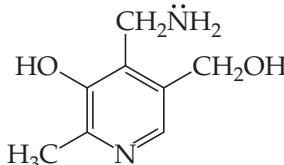
codeine
(3° amine)
heterocyclic amine



adrenalin
(2° amine)
aliphatic amine



benzocaine
(1° amine)
aromatic amine



pyridoxamine (vitamin B₆)
(1° and 3° amine)
aliphatic and aromatic amine

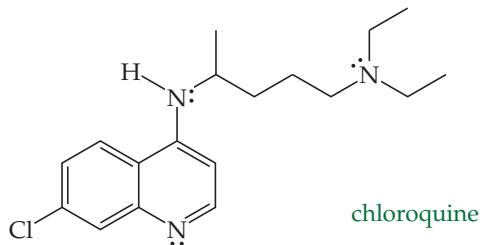
▲ **Figure 31.2** Industrially and biologically important amines. The lone pair on nitrogen has been added to highlight its position in these compounds.



Sample Exercise 31.1

Classifying Amines

Classify each amino group of chloroquine, a compound used for the treatment of malaria, as primary, secondary or tertiary. Comment also on whether they are aliphatic or aromatic amines.

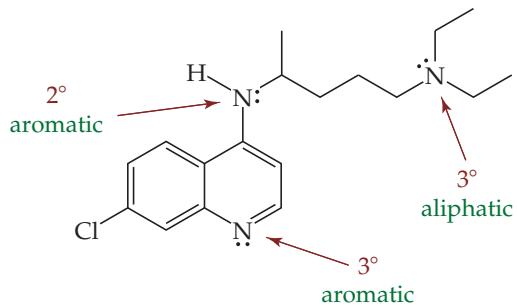


SOLUTION

Analyze We are asked to classify each amine in chloroquine.

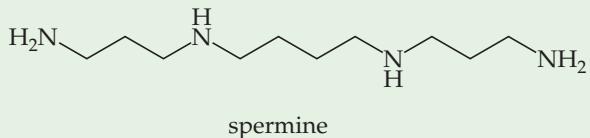
Plan Look for each nitrogen atom and determine if aromatic or aliphatic as a result of the nature of the bonding around them. Classify each as primary, secondary or tertiary by determining how many $\text{N}-\text{C}$ bonds exist.

Solve



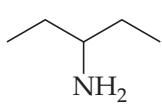
► Practice Exercise

Classify each amino group of spermine, a compound isolated from semen, as primary, secondary or tertiary.

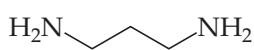


and pyridoxamine. Its role in the body is varied but essential for good health. Benzocaine is an effective local and topical anesthetic used in sting medication, sunburn cream and by dentists in mucosal tissue.

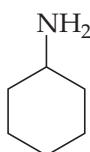
The systematic naming of simple monoaliphatic amines follows the same guidelines as for alcohols. The *e* of the parent alkane is replaced by *amine* (as opposed to *ol* for alcohols). Otherwise, the name is derived from the alkyl group attached to nitrogen and named as the alkylamine. For example:



3-pentanamine

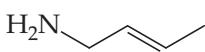


propane-1,3-diamine

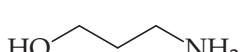


cyclohexylamine

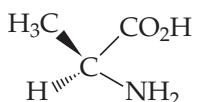
For amines containing multiple alkyl groups bonded to carbon, the amine is named based on those alkyl groups (Figure 31.1). Amines have a low priority in the nomenclature hierarchy. In many instances, the **amino group** is named as a substituent rather than as the class of an organic compound:



1-aminobut-2-ene



3-aminopropan-1-ol



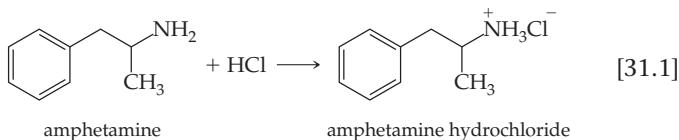
(*S*)-2-aminopropanoic acid

CHEMISTRY AND LIFE Amines and Amine Hydrochlorides

Many amines with low molecular weights have unpleasant “fishy” odors. Amines and ammonia (NH_3) are produced by the anaerobic (absence of O_2) decomposition of dead animal or plant matter. Two such amines with very disagreeable odors are $\text{H}_2\text{N}(\text{CH}_2)_4\text{NH}_2$, known as *putrescine*, and $\text{H}_2\text{N}(\text{CH}_2)_5\text{NH}_2$, known as *cadaverine*. Despite their usually unpleasant odors, amines and the products of their reactions are very important compounds in biological systems.

The pigment of vision, rhodopsin, is prepared by the formation of an imine bond between the biologically active aldehyde 11-*cis*-retinal and the protein opsin (Figure 31.3). When rhodopsin absorbs a photon of light, the *cis* double bond at C11 undergoes an isomerization to the *trans* geometry, which triggers a change in the shape of rhodopsin, leading to nerve impulses that are detected by the brain as a visual image.

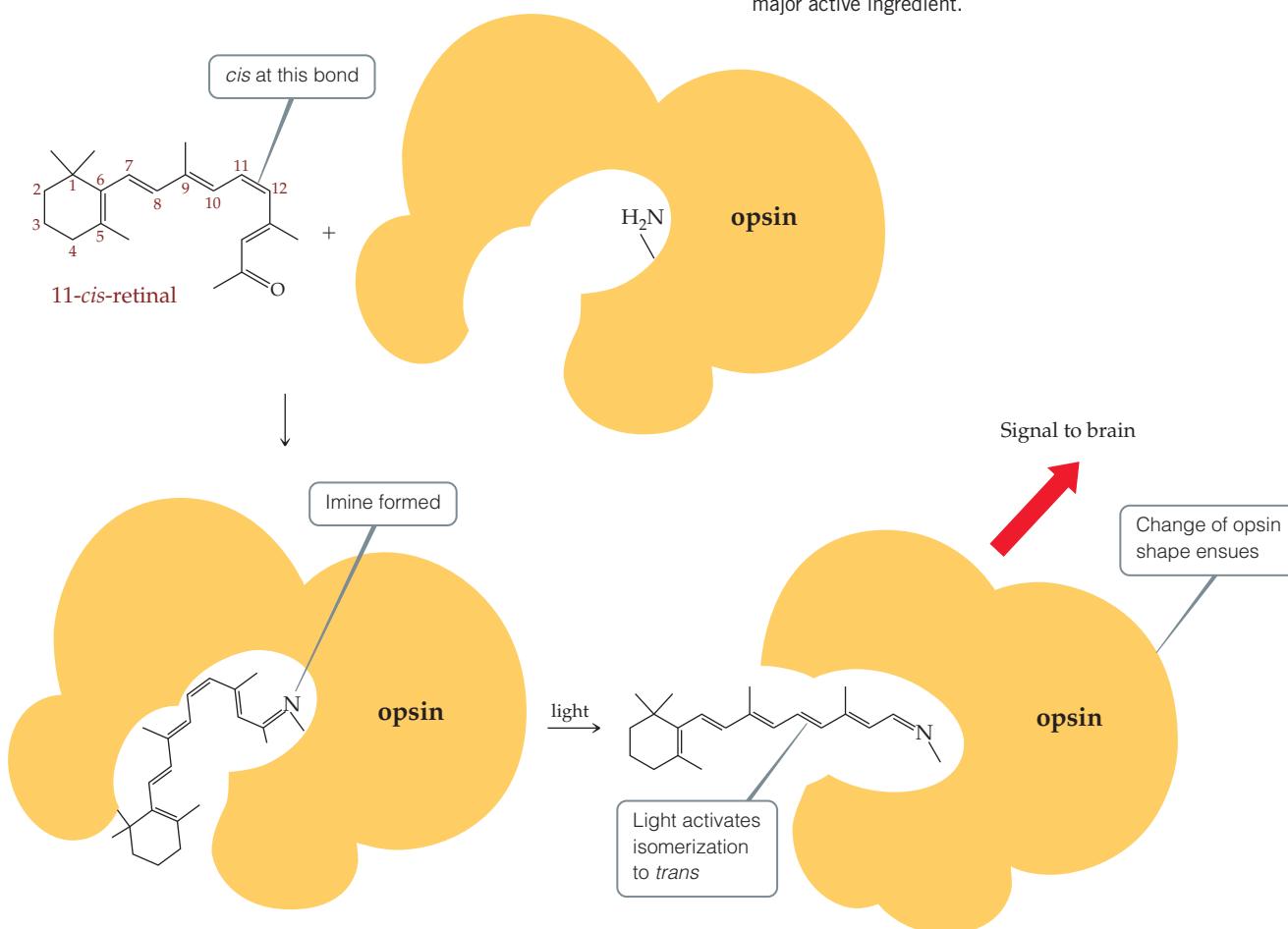
Many drugs, including quinine, codeine, caffeine and amphetamine (benzedrine), are amines. Like other amines, these substances are weak bases; the amine nitrogen is readily protonated by treatment with an acid. The resulting products are called ammonium salts. For example, amphetamine hydrochloride is the ammonium salt formed by treating amphetamine with HCl.



Such ammonium salts are less volatile, more stable and generally more water soluble than the corresponding neutral amines. Many drugs that are amines are sold and administered as ammonium salts for this very reason. Some examples of over-the-counter medications and cleaning products that contain amine hydrochlorides as active ingredients are shown in Figure 31.4.

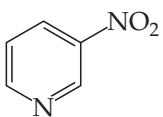


▲ Figure 31.4 Ammonium salts. Many commercial products, such as over-the-counter medications and antiseptics, contain amine hydrochlorides as the major active ingredient.

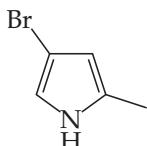


▲ Figure 31.3 Rhodopsin. Imine formation between 11-*cis*-retinal and the protein opsin is important for vision.

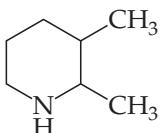
The names of heterocyclic amines are derived from the parent aliphatic or aromatic heterocycle; for example:



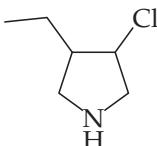
3-nitropyridine



4-bromo-2-methylpyrrole

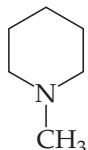


2,3-dimethylpiperidine

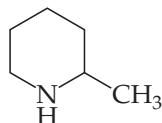


3-chloro-4-ethylpyrrolidine

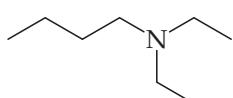
When more than one type of alkyl group (for example, methyl and ethyl groups) are bonded to nitrogen in the same molecule, the name must also reflect the position of the alkyl groups on nitrogen by using the prefix *N*. For example, the name *N*-methylpiperidine indicates that the methyl substituent is located on the nitrogen of the piperidine ring and not located on the carbon framework (for example, 2-methylpiperidine). The name *N,N*-diethyl-1-butanamine is used to describe both ethyl groups bonded to nitrogen. This compound differs from *N*-ethyl-2-ethyl-1-butanamine and 4-ethyl-3-hexanamine. All three compounds are constitutional isomers.

*N*-methylpiperidine

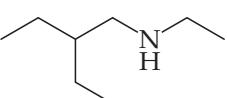
differs from



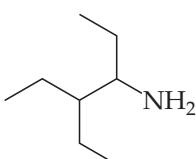
2-methylpiperidine

*N,N*-diethyl-1-butanamine

differs from

*N*-ethyl-2-ethyl-1-butanamine

differs from

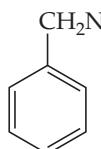


4-ethyl-3-hexanamine

Sample Exercise 31.2

Naming Amines

Write the systematic or IUPAC name for the following amines:



SOLUTION

Analyze We are asked to provide systematic names for three different amines

Plan Use the rules described earlier to structure the names.

Solve

(a) 1-pantanamine—derived from pentane.

(b) Benzylamine—derived from the benzyl group. This is better than naming the structure phenylmethanamine.

(c) Trimethylamine—derived from the three methyl substituents attached to nitrogen. Note that the prefix *N* is not required.

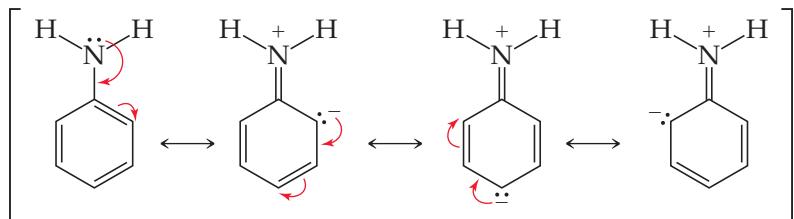
► Practice Exercise

Draw the structures for the following amines:

- (a) *tert*-butylamine,
- (b) 4-nitroaniline,
- (c) *N*-methyl-2-hydroxypiperidine

 Go Figure

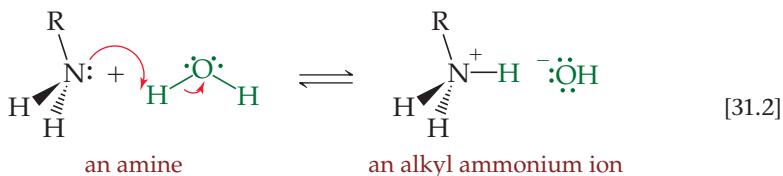
Why is cyclohexylamine a stronger base than aniline?



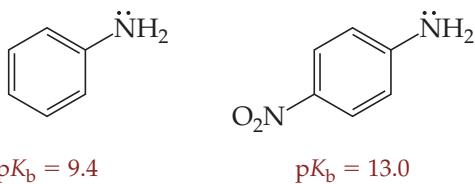
▲ Figure 31.5 Resonance contributors for aniline. The electron pair distribution responsible for the mild basicity and nucleophilicity in aromatic amines.

Reactivity of Amines

We introduced the nucleophilic nature of amines in Chapter 28. Along with this nucleophilicity is the ability of primary, secondary and tertiary amines to act as a weak base. The presence of the amine functional group allows for extensive solvation in aqueous media. This means that most amines have some solubility in water and hence can be tested with litmus paper to determine their presence. The acid-base reaction for a primary amine in aqueous solution can be written generally as:

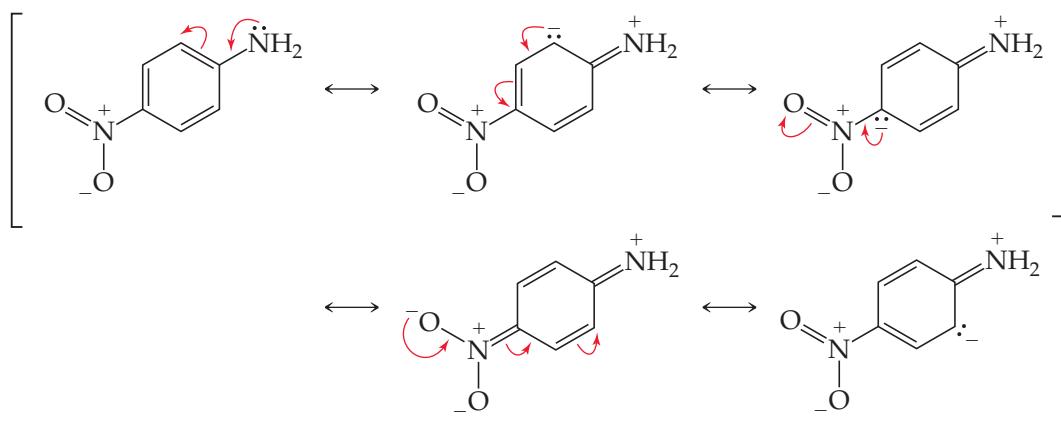


There are a few features to note in this reaction. First, the reaction is an equilibrium reaction, although this equilibrium lies to the left for all amines. Primary, secondary and tertiary alkylamines are over 10^6 times more basic ($pK_b = 3\text{--}4$) than comparable aromatic amines ($pK_b = 9\text{--}10$). The difference in basicity can be attributed to the resonance effects available to aromatic amines, such as shown for aniline in Figure 31.5. By distributing electrons around the ring, the nucleophilicity and basicity of aromatic amines is weakened compared with their aliphatic counterparts.

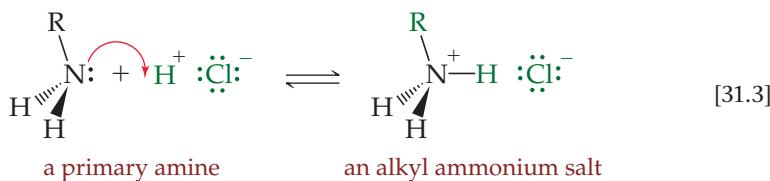


The basicity of aromatic amines is also influenced by the substituents on the aromatic ring. For example, 4-nitroaniline (**Figure 31.6**) is a much weaker base than aniline, due to resonance and inductive effects leading to a redistribution of electron density onto the NO_2 group.

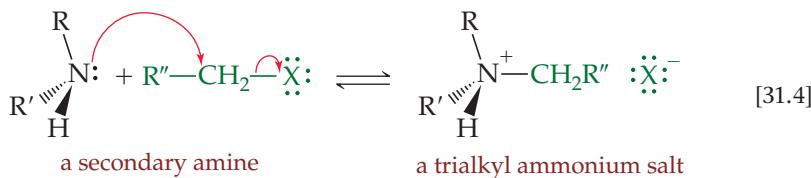
The second feature to note is that the acid-base reaction is written so that it involves the lone pair on nitrogen. The two electrons in this lone pair form a new N-H covalent bond. Finally, the product of the reaction is an **alkyl ammonium salt**, also known as a 4° ammonium salt. Such ammonium salts are easily formed in essentially quantitative yield by reaction of an amine with strong acid such as aqueous HCl:



▲ Figure 31.6 Rationalising the basicity of 4-nitroaniline. There are more resonance contributors to the structure of 4-nitroaniline when compared to aniline, suggesting the basicity (and nucleophilicity) of the substituted aniline is less. This situation is accentuated by the electron-withdrawing effect of the nitro group, which is inductive throughout the molecule.



Such salts are usually highly soluble in water. Alkyl ammonium salts can also be formed by the reaction of amines with haloalkanes. In the following example, a secondary amine is reacted with a haloalkane to form an alkyl ammonium salt:



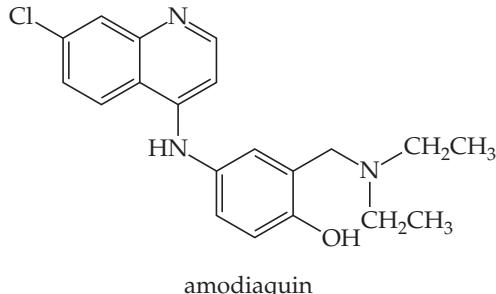
The alkylation reaction is general but care needs to be taken in stoichiometry of reagents, choice of solvent and choice of haloalkanes in order to limit further alkylation.



Sample Exercise 31.3

Amine Basicity

Amodiaquin is commonly used to treat quinine-resistant malaria. Identify the amine groups within this compound and indicate which of the nitrogen atoms of amodiaquin is the most basic.



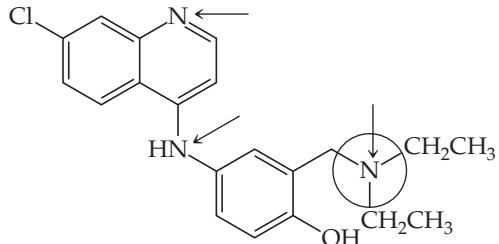
Continued

SOLUTION

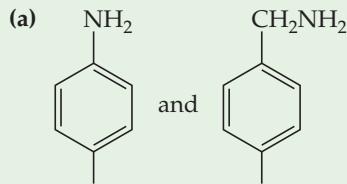
Analyze We are asked to classify the amines located within the structure of amodiaquin and from there, which is the most basic.

Plan Identify the amines and look for aliphatic amines as the most basic.

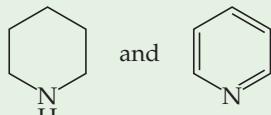
Solve The comparison here is between two aromatic amines and an aliphatic amine (all amine groups indicated by an arrow). The aliphatic amine (circled) is most basic.

**► Practice Exercise**

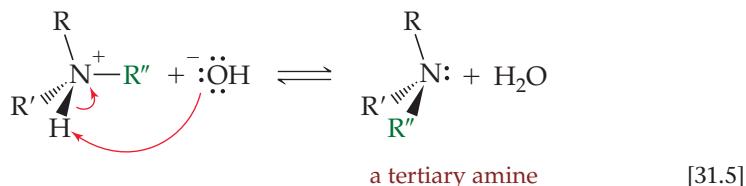
Select the stronger base from each amine pair.



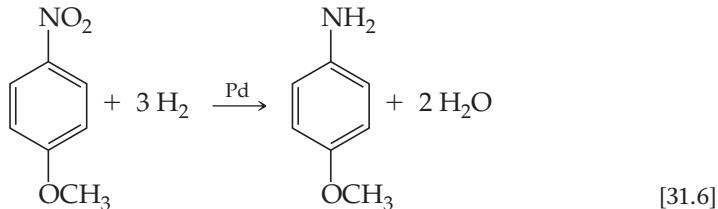
(b)

**Synthesis of Amines**

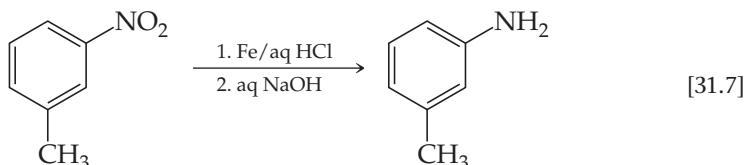
Amines can be formed in several ways, some simple and some complex. We look at three of the most common ways to form amines. Equation 31.4 illustrates the nucleophilicity of amines and their reactivity towards alkyl halides. The ammonium ion formed in the reaction is a weak acid ($pK_a = 9\text{--}11$), comparable in pK_a to phenols. Addition of an aqueous base during workup (see Equation 31.5) leads in this case to the tertiary amine. Hence, by a simple *alkylation* we have chemically transformed a secondary amine into a tertiary amine.

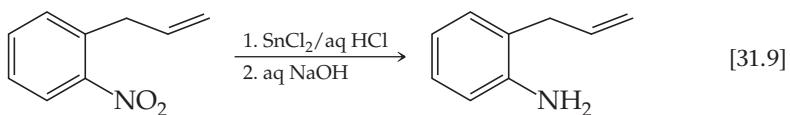
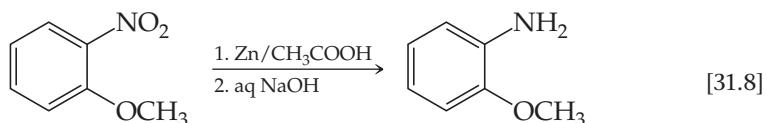


Aromatic amines such as 4-methoxyaniline are easily prepared by reduction of the corresponding nitro compound. This reduction is usually achieved by catalytic hydrogenation—that is, using H_2 in the presence of a metal catalyst such as Ni, Pd or Pt. Although this reaction is very general, easily accomplished and high yielding, care needs to be taken to ensure that other groups susceptible to hydrogenation, such as double bonds, are not also reduced (recall that benzene does not react like an alkene towards addition type reactions).

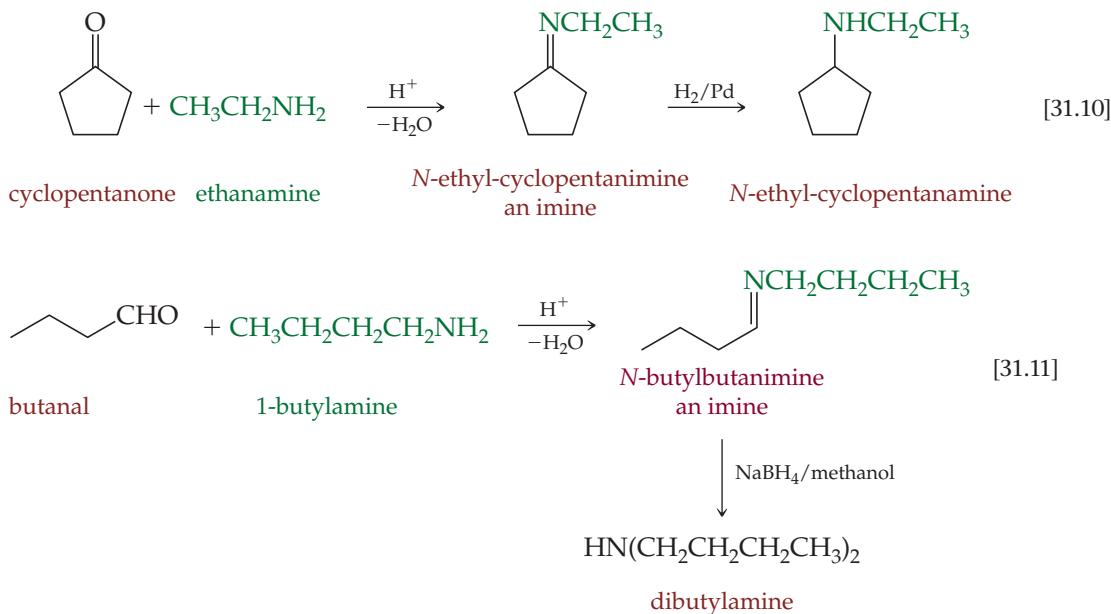


A metal reduction can also be used to effect the same transformation. Metals such as Fe and Zn, and metal salts such as $SnCl_2$, can be used. For example, reduction of the nitro group in Equation 31.6 by hydrogenation (H_2/Pd) would also reduce the alkene group. The use of tin(II) chloride as a reducing agent is more selective for the nitro group, yielding the desired product.





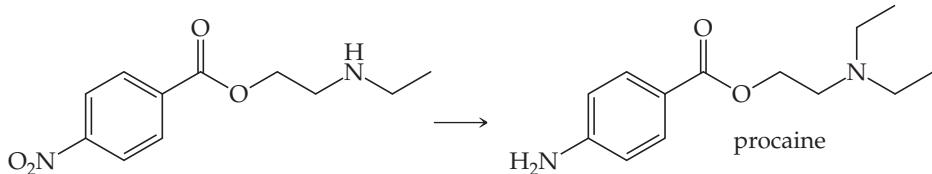
Recall that imines are formed by the reaction of aldehydes and ketones with primary amines. The reactivity of the imine group resembles that of a carbonyl group, meaning that both functional groups are susceptible to hydride reduction and catalytic hydrogenation. This method, known as a **reductive amination**, is an easy way of producing secondary amines from primary amines.



Sample Exercise 31.4

Amine Synthesis

Describe a synthesis for procaine, a local anaesthetic, from the suggested starting material. You may choose any reagents and conditions you like that will affect the transformations needed.



SOLUTION

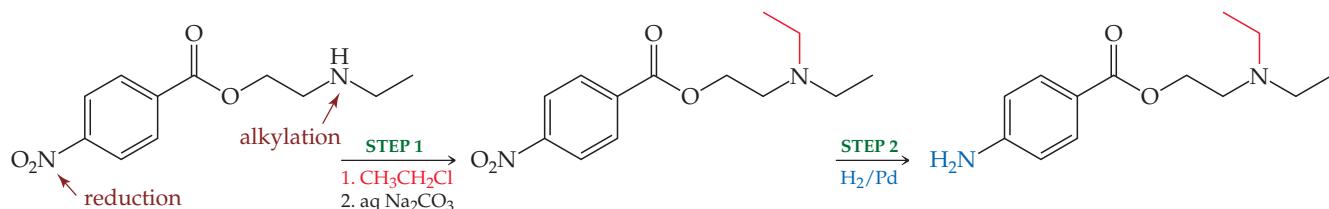
Analyze We are given a compound and asked to devise a synthesis of procaine.

Plan Identify the functional group differences between starting material and product and then a plan of attack to make the necessary interconversions.

Solve In this case, the differences in functional groups can be achieved by an alkylation, introducing the ethyl group, and a reduction of the nitro group to give the amine. *Be very careful with the order of reaction.* In this example, alkylation must occur before reduction in order to yield procaine. Reducing the nitro group to the amino group, followed by alkylation, would be likely to cause a reaction to occur on both nitrogens. The difference in

Continued

nucleophilicity is not significant enough to cause alkylation of just the aliphatic amine.

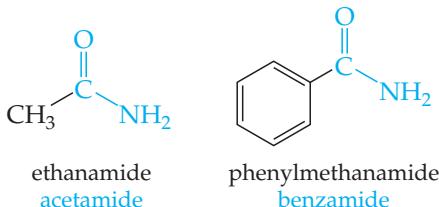


► Practice Exercise

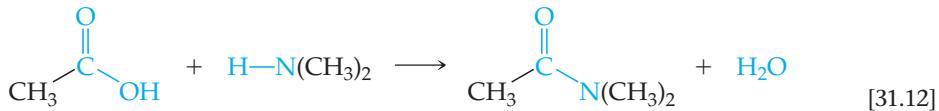
Describe a method for separating a mixture of nitrobenzene from aniline in diethyl ether solution.

Amides

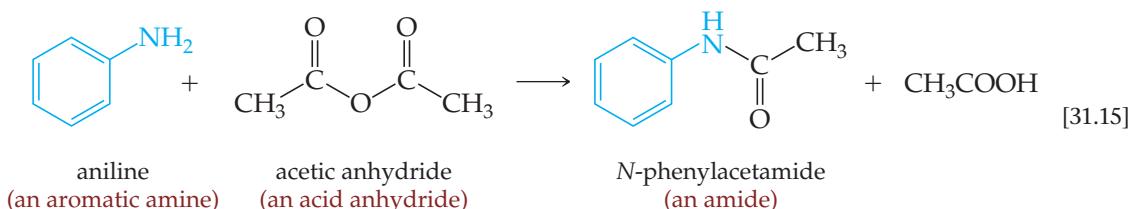
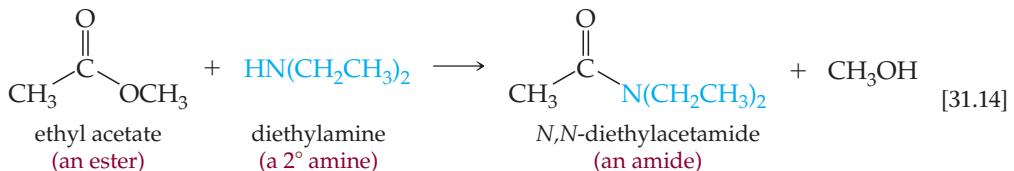
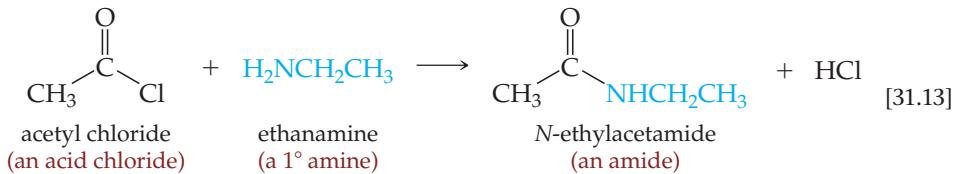
Amides are derivatives of carboxylic acids in which the OH group of the carboxylic acid group has been replaced with a NRR' group ($\text{R}, \text{R}' = \text{H, alkyl, aryl}$), as in these examples:



An amide can be considered as the condensation product of an amine and a carboxylic acid:

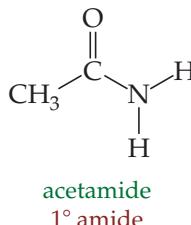


In Section 29.5, we saw that amides can also be formed by reaction of an acid anhydride, acid chloride or ester with ammonia, a primary or secondary amine. Examples of such reactions are:



Amides are classified as primary, secondary or tertiary, depending on their substitution at the amide nitrogen.

- **Primary (1°) amide:** the nitrogen of the amide group is bonded to one carbon atom, which is that of the carbonyl group. For example:

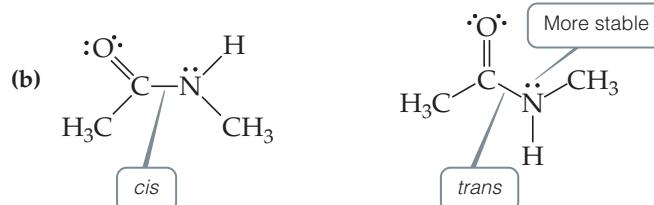
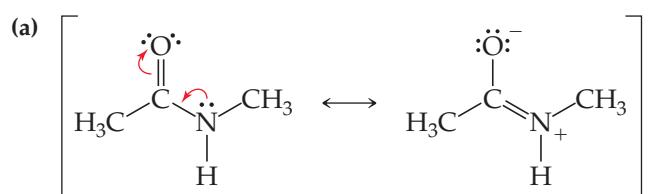
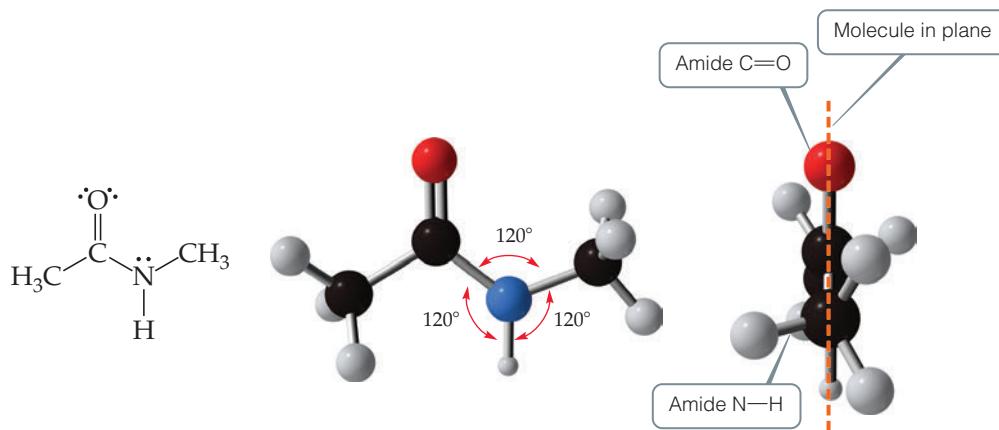


- **Secondary (2°) amide:** the nitrogen of the amide group is bonded to two carbon atoms, one of which is a carbonyl group; for example *N*-phenylacetamide.
- **Tertiary (3°) amide:** the nitrogen of the amide group is bonded to three carbon atoms, one of which is a carbonyl group. The other two carbon atoms usually belong to alkyl or aryl groups; for example *N,N*-diethylacetamide.

The geometry of the amide bond is quite unexpected. At first glance, you would expect the bonding geometry about the nitrogen atom to be a trigonal pyramid, similar to what you would find for an amine. During the 1930s Linus Pauling discovered that an amide is planar through the CONR group (Figure 31.7), commensurate with the observed bond angles of approximately 120° about the amide nitrogen.

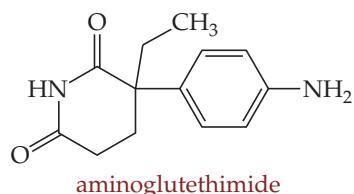
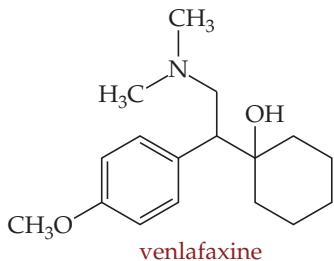
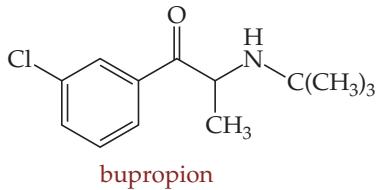
Pauling rationalized the unusual geometry of the amide bond by describing two resonance hybrid structures (Figure 31.8(a)). One structure has a single bond between the carbonyl carbon and nitrogen, whereas the other has a carbon-nitrogen double bond. The overall structure is neither of these, but a hybrid containing a significant amount of carbon-nitrogen double-bond character. It is this double-bond character that makes the six

◀ **Figure 31.7** The geometry of the amide bond. Bond angles of approximately 120° are seen about the amide nitrogen. This trigonal-planar geometry is coplanar with the amide carbonyl group.

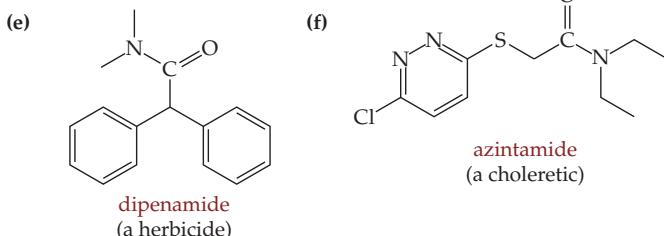
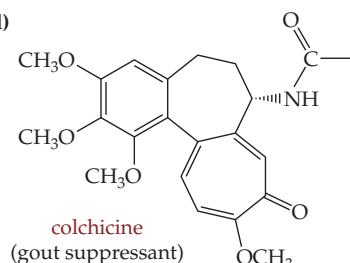
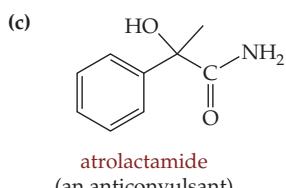
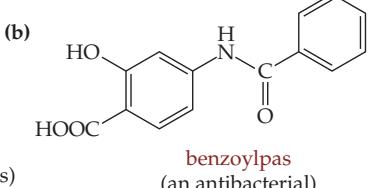
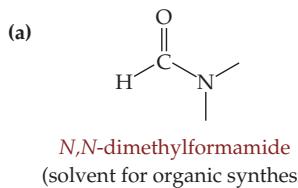


◀ **Figure 31.8** Conformation and rationalization. (a) The planar nature of the amide bond can be rationalized by resonance delocalization. (b) Amides can exist as *cis* or *trans* conformers known as rotamers because there is limited rotation about the C—N bond. The *trans* conformer is favoured because it reduces any steric strain caused by substituents.

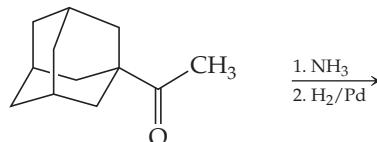
- 31.7** *Bupropion* is an antidepressant drug also used to treat attention-deficit hyperactivity disorders. *Venlafaxine* (sold under the brand name Effexor®) is an antidepressant drug. *Aminoglutethimide* has been used as an anticonvulsant. Classify bupropion, venlafaxine and aminoglutethimide as primary, secondary or tertiary amines.



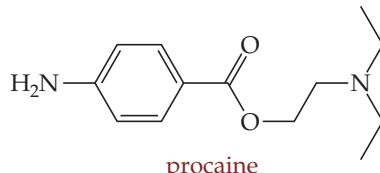
- 31.8** Define the following amides as primary, secondary or tertiary:



- 31.9** *Rimantadine* is effective in preventing infections caused by the influenza A virus. It can be formed from the ketone and the reagents shown. Suggest a structure for rimantadine consistent with the reaction conditions. [Hint: The reaction is undertaken in two steps.]

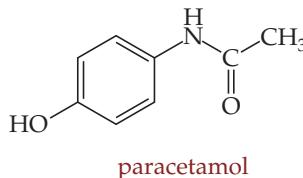


- 31.10** *Procaine* was one of the first local anaesthetics used. Its hydrochloride salt is marketed as *Novocaine*®.



- (a) Draw the product of the reaction of procaine with 1 mole equivalent of HCl. (b) Suggest a synthesis for procaine starting with 4-nitrobenzoic acid.

- 31.11** Devise a synthesis for *paracetamol* beginning with 4-nitrophenol.



31.1 (b) 31.2 (c) 31.3 (b)

Answers to Self-Assessment Exercises

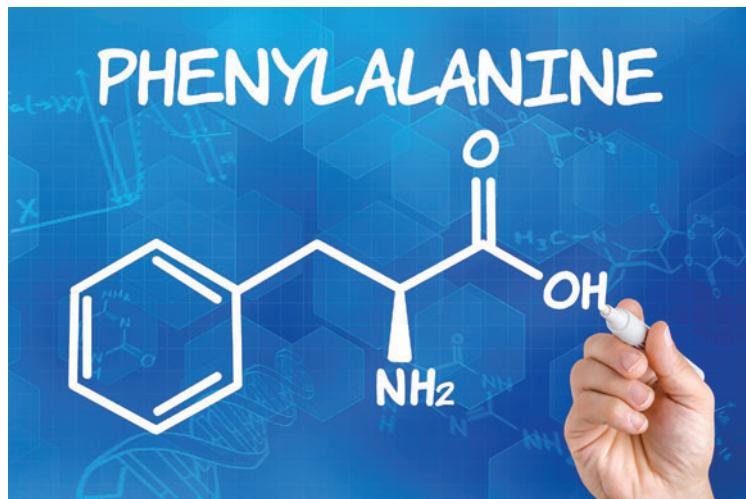
31.1 (b)

31.2 (c)

31.3 (b)



31.2 | Amino Acids



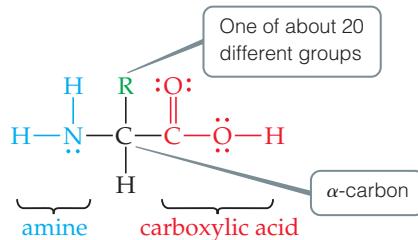
Phenylalanine is one of the two amino acids that make up the sweetener aspartame, an artificial sweetener sold under the trade names “Equal®” and “Nutrasweet®”. Phenylalanine also happens to be one of the essential amino acids, which means that we must have it in our diets to survive long term. However, in about one out of 10,000 to 20,000 Caucasian or Asian births, the enzyme that converts phenylalanine to another amino acid, tyrosine, is completely or nearly completely absent because of a genetic defect. The result is that phenylalanine accumulates in the blood and in body tissues. The disease that results is called phenylketonuria (PKU), which causes intellectual disability and seizures. This is why cans of diet soft drink sometimes carry the warning “Phenylketonurics: contains phenylalanine”. Newborns are routinely tested for PKU when they are about three days old.

Although biological systems are unimaginably complex, they are nevertheless constructed of molecules of quite modest size. The example here illustrates that, in order to understand biology, we need to understand the chemical behavior of molecules with low molar mass, as well as much larger molecules, which form the basis of *biological chemistry*.

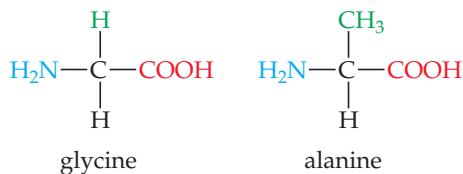
By the end of this section, you should be able to:

- understand the acid-base properties of an amino acid
- know how an amino acid can be made in the laboratory

Amino acids contain both an amino group and a carboxyl group. An **α -amino acid** is an amino acid in which the amino group is bonded to the carbon adjacent to the carboxyl group. α -Amino acids play a large role in biological chemistry as the building blocks of protein, and their properties are dictated somewhat by the nature of the R group.



In *glycine*, which is the simplest amino acid, R is a hydrogen atom; in *alanine*, R is a methyl group.



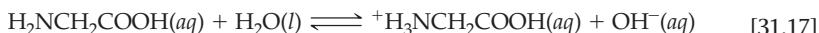
Because amino acids contain a carboxyl group, they can serve as weak acids. They also contain the NH_2 group, characteristic of amines, and thus they can also act as weak bases.

Amino acids, therefore, are **amphiprotic**. For glycine, we might expect that the acid and the base reactions with water would be as follows:

Acid:

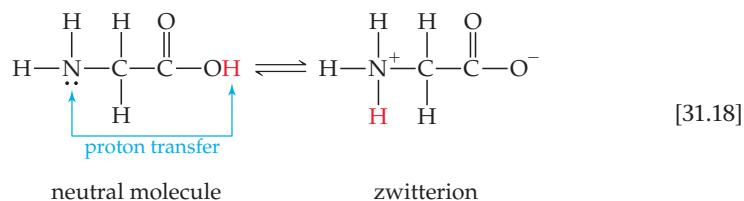


Base:



The pH of a solution of glycine in water is about 6.0, indicating that it is a slightly stronger acid than a base.

The acid-base chemistry of amino acids is somewhat more complicated than shown in Equations 31.16 and 31.17, however. As a result of the fact that the COOH group can act as an acid and the NH₂ group can act as a base, amino acids undergo a “self-contained” Brønsted-Lowry acid-base reaction in which the proton of the carboxyl group is transferred to the basic nitrogen atom:



Although the form of the amino acid on the right side of Equation 31.18 is electrically neutral overall, it has a positively charged end and a negatively charged end. A molecule of this type is called a **zwitterion** (German for “hybrid ion”). This doubly ionized form predominates at near-neutral values of pH. Amino acids are more properly written as the zwitterion because of this fact, although there are many instances in science (not necessarily chemistry) where the un-ionized form (left side of Equation 31.18) is commonly written.

Do amino acids exhibit any properties indicating that they behave as zwitterions? If so, they should behave in a similar manner to ionic substances. Crystalline amino acids (Figure 31.10) have relatively high melting points, usually above 200 °C, which is characteristic of ionic solids. Amino acids are far more soluble in water than in non-polar solvents. In addition, the dipole moments of amino acids are large, consistent with a large separation of charge in the molecule. Thus the ability of amino acids to act simultaneously as acids and bases has important effects on their properties.

With the exception of glycine, all naturally occurring amino acids have at least one stereocenter, the α -carbon. Figure 31.11 shows the two enantiomers of the amino acid alanine. For historical reasons, the two enantiomers are distinguished by the labels D and L. All amino acids normally found in proteins have the L configuration at the stereocenter (except for glycine, which is not chiral). In terms of absolute stereochemistry, all naturally occurring α -amino acids have the S configuration (Section 25.3). This means that the enantiomer or D-amino acid has R configuration.

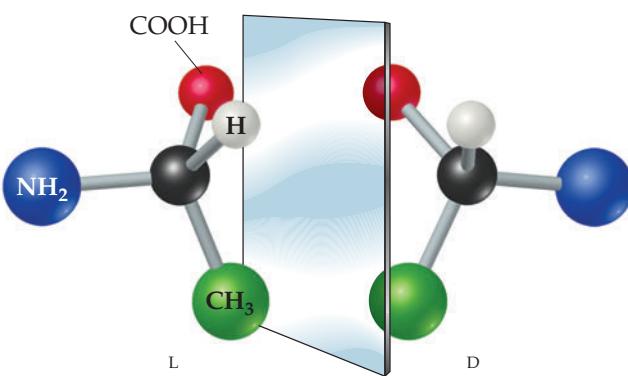
Figure 31.12 shows the structural formulas of the 20 most common α -amino acids found in proteins. Our bodies can synthesize 10 of these α -amino acids in sufficient



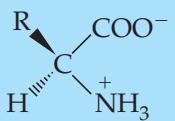
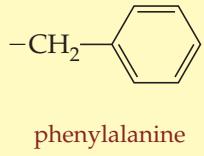
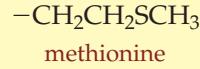
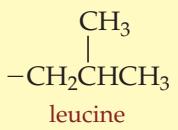
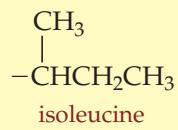
▲ **Figure 31.10 Lysine.** One of the amino acids found in proteins, lysine is available as a dietary supplement. The L in the name L-Lysine refers to a specific configuration of atoms that is found in naturally occurring amino acids.

Go Figure

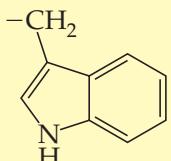
Are all α -amino acids chiral?



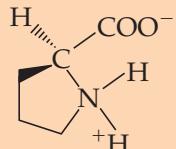
▲ **Figure 31.11 Alanine enantiomers.** The middle carbon of alanine is stereogenic, and therefore there are two enantiomers, which by definition are non-superimposable mirror images of each other.

L-amino acid general structure**Non-polar side chains (R=)**

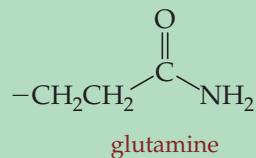
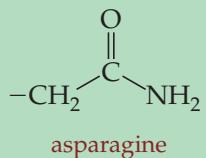
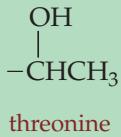
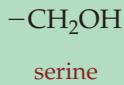
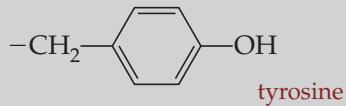
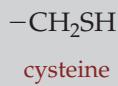
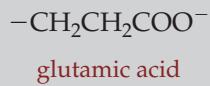
phenylalanine



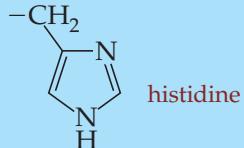
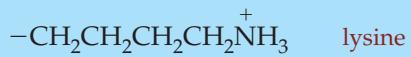
tryptophan



proline

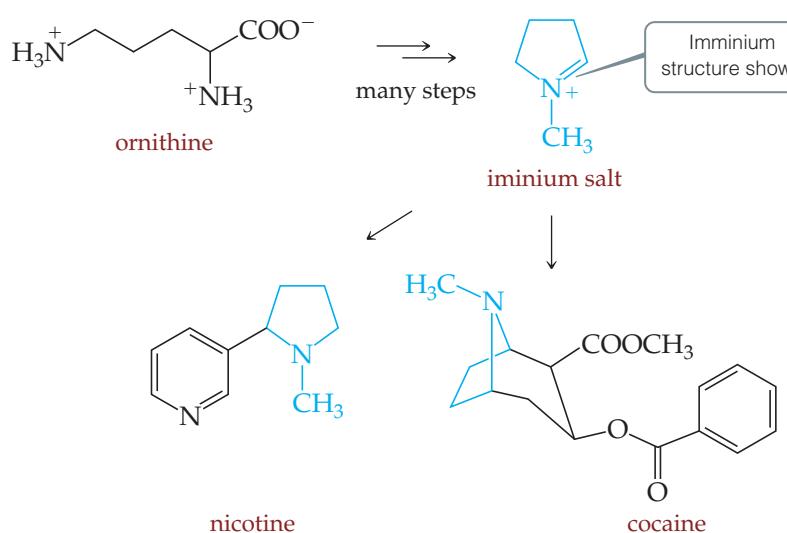
Polar side chains (R=)**Acidic side chains (R=)**

tyrosine

Basic side chains (R=)

histidine

▲ **Figure 31.12** The 20 common α -amino acids found in the human body. Each ionizable group within the acidic and basic classes is shown in the form present in highest concentration at pH = 7.0.



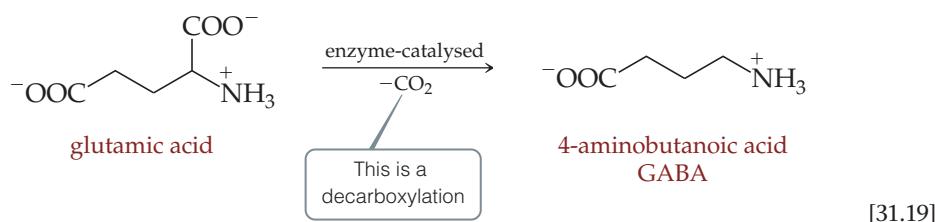
◀ **Figure 31.13** Ornithine. This amino acid is a relative of lysine but with a shorter side chain.

amounts for our needs. The other 10 must be ingested and are called *essential amino acids* because they are necessary components of our diet.

α -Amino acids, which make up the 20 most common naturally occurring amino acids, can be classified by the nature of their **side chain**, which is bonded to the amino acid's α -carbon. These side chains are used to classify amino acids as polar and non-polar. The polar amino acids can be further classified as ionized (for example, containing acidic and basic side chains) and un-ionized (that is, neutral-polar side chains). Proline is a cyclic amino acid, which differs from the acyclic nature of the other 19 amino acids.

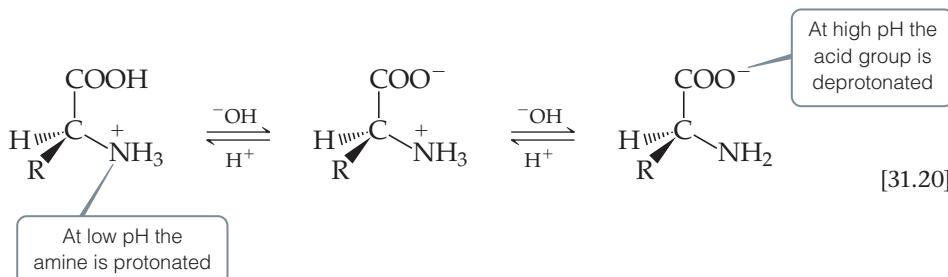
Although all amino acids are chiral, with the exception of glycine, isoleucine and threonine also contain a second stereocenter. Four stereoisomers are possible, but only *one* of the four stereoisomers is found in proteins.

Two amino acids not listed in Figure 31.12 deserve special mention because of their social and medicinal importance. Ornithine is an amino acid found in the liver, with importance in the urea cycle. It is also a precursor for the biosynthesis of nicotine and tropane alkaloids such as cocaine (Figure 31.13). GABA (γ -aminobutyric acid) is a neurotransmitter (a substance that transmits nerve impulses across a synapse), found exclusively in the brain and central nervous system (CNS). It has been implicated in several neurodegenerative conditions, including memory loss. GABA is formed within the CNS by the enzyme-catalyzed decarboxylation of glutamic acid. Its systematic name, 4-aminobutanoic acid, is less commonly used than its acronym, GABA.



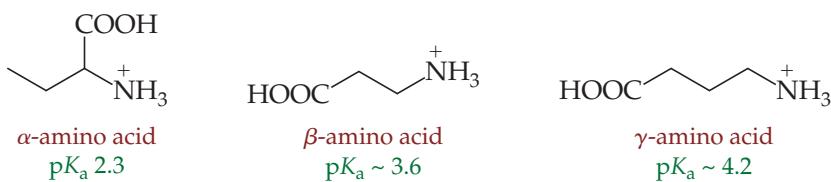
Acid–Base Properties

The amphiprotic nature of α -amino acids can be described by the following equation:



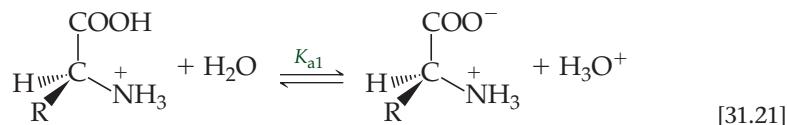
► Figure 31.14 Carboxyl group acidity.

The inductive effect of the ammonium ion changes the pK_a of the COOH of amino acids. The effect of the ammonium group is connectivity dependent and approaches that of a normal carboxylic acid as the number of intervening bonds increase.



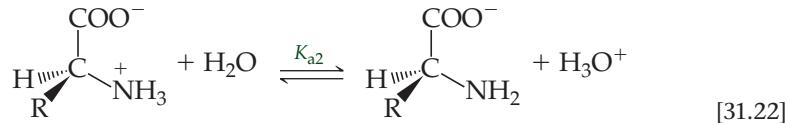
At low pH the basic amino group is protonated, forming the ammonium ion. In neutral solution the zwitterion is formed. The pH at which the zwitterion concentration is the highest is called the **isoelectric point**. In basic solution, the carboxylic acid group is ionized to form the carboxylate group. Notice that amino acids are ionic at all pHs, which makes them highly water soluble and difficult to purify by the normal procedures used in organic chemistry.

Consider the following reaction:



The average pK_a value (pK_{a1}) for the carboxyl group of an α -amino acid is ~ 2 . This value is considerably lower than that for acetic acid ($pK_a = 4.76$), which implies that α -amino acids are stronger acids than acetic acid. The reason for this difference in pK_a can be attributed to the inductive electron-withdrawing effect of the ammonium group, which draws electron density out of the O—H bond, making it weaker relative to the same bond in acetic acid. **Figure 31.14** shows how this inductive effect weakens as the number of intervening bonds increases.

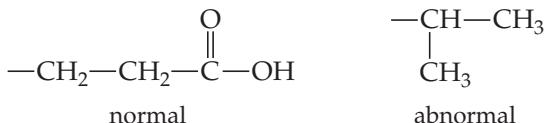
The ammonium group of an α -amino acid is a stronger acid than a comparable aliphatic ammonium ion due to the weak inductive effect of the carboxyl group. The average pK_a for an α -amino acid (pK_{a2}) is ~ 9.5 , compared with $pK_a \sim 10.5$ for an aliphatic amine. This implies that the amine of an α -amino acid is generally a weaker base than a 1° aliphatic amine.



A CLOSER LOOK Sickle-Cell Anemia

Our blood contains a complex protein called hemoglobin, which carries oxygen from the lungs to other parts of the body. In the genetic disease known as sickle-cell anemia, hemoglobin molecules are abnormal and have a lower solubility, especially in their unoxygenated form. Consequently, as much as 85% of the hemoglobin in red blood cells crystallizes from solution.

The reason for the insolubility of hemoglobin in sickle-cell anemia can be traced to a structural change in one part of an amino acid side chain. Normal hemoglobin molecules contain glutamic acid, which contributes to the solubility of the hemoglobin molecule in water. In the hemoglobin molecules of people suffering from sickle-cell anemia, the glutamic acid residue is replaced by valine:

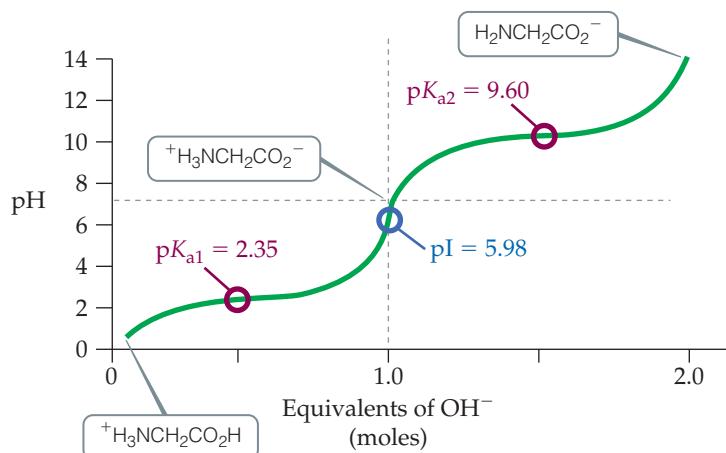


Valine is non-polar (hydrophobic), and its presence leads to the aggregation of the defective form of hemoglobin into particles too large to remain suspended in biological fluids. It also causes the cells to distort

into a sickle shape, as shown in **Figure 31.15**. The sickled cells tend to clog the capillaries, causing severe pain, physical weakness and gradual deterioration of the vital organs. The disease is hereditary and, if both parents carry the defective genes, it is likely that their children will possess only abnormal hemoglobin.



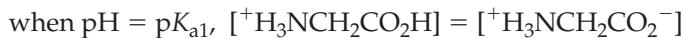
▲ **Figure 31.15** Normal and sickled red blood cells. Normal red blood cells are about 1 μm in diameter.



◀ **Figure 31.16** Titration of glycine with aqueous NaOH solution. pK_{a1} relates to the pK_a of the carboxylic acid, while pK_{a2} relates to the pK_a of the ammonium ion.

Let us now investigate what the pK_a value of an α -amino acid means and what information it provides. To do this, we need to consider the pH titration of a simple amino acid, such as glycine, with aqueous sodium hydroxide solution (Figure 31.16).

At low pH, glycine is in its fully protonated form, $^+H_3NCH_2CO_2H$. The addition of OH^- sees a gradual rise in pH until a plateau is reached (also described as a *point of inflection* on the titration curve). At this point, which corresponds to the addition of 0.5 mol of OH^- , the $pH = pK_{a1}$. This is the point at which the concentration of the zwitterion equals that of the ammonium ion:



A pH value *lower* than pK_{a1} means that the ammonium ion predominates. A pH value *higher* than pK_{a1} means the zwitterion predominates in solution.

The endpoint of this first reaction is reached when 1 mole equivalent of OH^- is added. This point, known as the *isoelectric point*, is where the zwitterion $^+H_3NCH_2CO_2^-$ is at its maximum concentration. At the isoelectric point, pI , the amino acid has no net charge. The pI of an amino acid usually lies at the midpoint of pK_{a1} and pK_{a2} . We can represent this mathematically by:

$$pI = \frac{1}{2}(pK_{a1} + pK_{a2}) \quad [31.23]$$

$$\text{for glycine, } pI = \frac{1}{2}(2.35 + 9.60) = 5.98$$

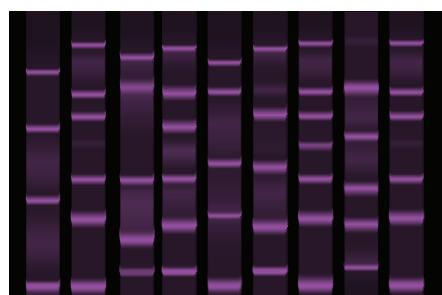
The addition of another 0.5 molar equivalents of OH^- causes the pH to rise, then plateau. This point of inflection indicates pK_{a2} , the pK_a value associated with the ammonium group. This is the point at which the concentration of the zwitterion equals that of the carboxylate ion:



A value of pH *lower* than pK_{a2} means that the zwitterion ion predominates. A value for the pH *higher* than pK_{a2} means that the amine and carboxylate groups predominate in solution.

The net charge of an amino acid is easily estimated using the pI value. As pH values fall below pI , the amino acid exists more and more as the ammonium ion, which is positively charged. At pH values higher than the pI , the amino acid exists as the carboxylate ion, which is negatively charged.

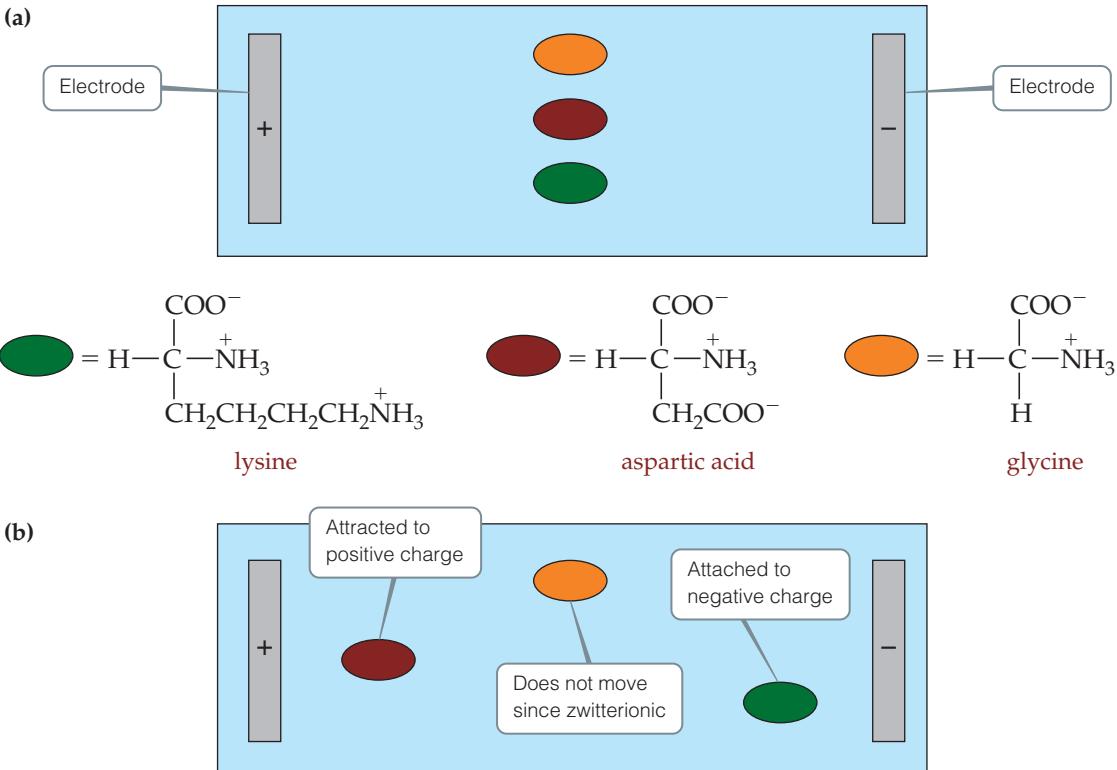
Knowledge about the pI of an amino acid is also useful for purification by **electrophoresis**, a process used to separate molecules based on charge (Figure 31.17). This useful technique uses either paper or gel as a separating medium and electric charge to separate amino acids, proteins and other charged molecules (Figure 31.18). You will learn more about this technique in courses devoted to biochemistry and cellular biology, where the separation of amino acids (and proteins) by this technique is a fundamental tool in the laboratory.



▲ **Figure 31.17** A typical electrophoresis result. The bands highlighted under UV light represent different amino acids, peptides or proteins that have been separated in a gel matrix. Each species can be isolated and then identified.

Go Figure

How would tyrosine, $pI = 5.66$, act at pH 5.66 in electrophoresis? Would this change at pH 9?



▲ **Figure 31.18 Separation of amino acids by electrophoresis.** By running an electrophoresis experiment at pH 5.98, we can separate three amino acids, based on their pI values. At pH 5.98, glycine has no net overall charge because of its zwitterionic state. Aspartic acid ($pI = 2.77$) is deprotonated at pH 5.98 and is attracted to the positive electrode. Lysine ($pI = 9.74$) is protonated at pH 5.98 and is attracted to the negative electrode. (a) At time = 0, (b) at time > 0.

Sample Exercise 31.5**Acid–Base Properties of Amino Acids**

Draw structural formulas for the predominant form of glycine (pK_a 2.35, 9.60) in aqueous solution at pH 1, 6 and 12.

SOLUTION

Analyze We are asked to draw the form of an amino acid at a given pH knowing the pK_a .

Plan When working through a problem like this, it is important to identify the relationship between the pH and the closest pK_a value.

Solve At pH 1, glycine is predominantly in the fully protonated form:



The reason is that $\text{pH} < pK_{a1}$.

pH 6 is between the two given pK_a values, so the zwitterion is the major form:



At pH 12, the carboxylate form predominates:



The reason is that $\text{pH} > pK_{a2}$.

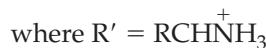
To verify the answer to this problem, you could use Figure 31.16.

► Practice Exercise

Draw structural formulas for the predominant forms of serine (pK_a 2.21, 9.15) in aqueous solution at pH 2, 7 and 10.

We can predict which species predominates at a certain pH in a much more quantitative way. For example, the equation for the acid-dissociation constant K_{a1} , shown in Equation 31.21, can be written as follows:

$$K_{a1} = \frac{[\text{H}_3\text{O}^+][\text{R}'-\text{CO}_2^-]}{[\text{R}'-\text{COOH}]} \quad [31.24]$$



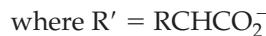
Rearrangement gives

$$\frac{K_{a1}}{[\text{H}_3\text{O}^+]} = \frac{[\text{R}'-\text{CO}_2^-]}{[\text{R}'-\text{COOH}]} \quad [31.25]$$

Knowing the value of $\text{p}K_{a1}$ and the pH of the solution (which determines the value of $[\text{H}_3\text{O}^+]$), we can calculate the exact ratio of the fully protonated and zwitterion forms at values between pH = 0 and the pI, as well as the net charge on the α -carboxyl group at that pH.

In a similar way, the equation for the acid-dissociation constant K_{a2} , shown in Equation 31.22, can be written as follows:

$$K_{a2} = \frac{[\text{H}_3\text{O}^+][\text{R}'-\text{NH}_2]}{[\text{R}'-\text{NH}_3^+]} \quad [31.26]$$



Rearrangement gives

$$\frac{K_{a2}}{[\text{H}_3\text{O}^+]} = \frac{[\text{R}'-\text{NH}_2]}{[\text{R}'-\text{NH}_3^+]} \quad [31.27]$$

Knowing the value of $\text{p}K_{a2}$ and the pH of the solution, we can calculate the exact ratio of the amine and zwitterion forms at values between pI and pH = 14, as well as the net charge on the α -amino group at that pH.

These calculations are similar to those used for buffer solutions using the Henderson–Hasselbalch equation for weak acid/conjugate base pairs:

$$\text{pH} = \text{p}K_a + \log \frac{[\text{conjugate base}]}{[\text{weak acid}]} \quad [31.28]$$

Sample Exercise 31.6

Estimating Net Charge

Let's reconsider the question posed in Practice Exercise 31.5. Estimate the net charge on serine at pH = 2.

SOLUTION

gives

Analyze We are asked to determine the percentage of each species in solution to determine the net charge at a given pH.

Plan Use the Henderson–Hasselbalch equation to determine the ratio of acid to conjugate base, hence the percentage of each in solution.

Solve Using Equation 31.28 and substituting in values:

$$\text{pH} = 2 \text{ and } \text{p}K_a = 2.21$$

$$2 = 2.21 + \log \frac{[\text{R}'\text{CO}_2^-]}{[\text{R}'\text{CO}_2\text{H}]}$$

$$\frac{[\text{R}'\text{CO}_2^-]}{[\text{R}'\text{CO}_2\text{H}]} = 0.62$$

Continued

This ratio concludes that the amino acid exists as 38% $\text{R}'\text{CO}_2^-$ and can be calculated using:

$$\frac{x}{100-x} = 0.62$$

where x = the percentage of $\text{R}'\text{CO}_2^-$

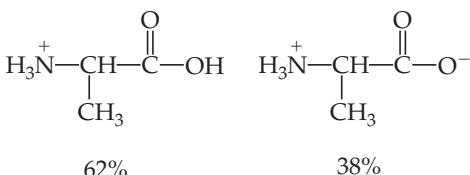
A similar calculation for the amino group gives

$$2 = 9.15 + \log \frac{[\text{R}'\text{NH}_2]}{[\text{R}'\text{NH}_3^+]}$$

$$\frac{[\text{R}'\text{NH}_2]}{[\text{R}'\text{NH}_3^+]} = 7.1 \times 10^{-8}$$

which indicates that the amino group is $> 99.9\%$ protonated at pH 2.

The average charge of the amino acid in solution can be calculated from the contributions of each charged species at pH 2. That is, at pH 2 the following two species exist:



Their respective charges are +1 and 0.

Therefore, based on their relative proportions in solution,

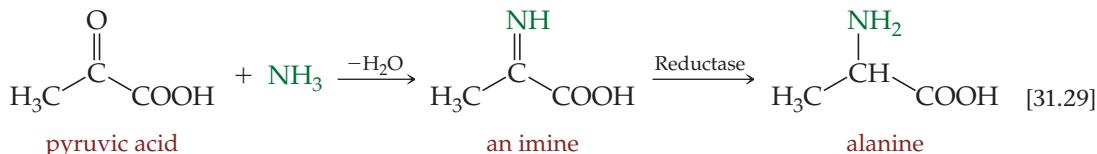
$$\text{Average charge} = 0.62 \times 1.0 + 0.38 \times 0 = +0.62$$

► Practice Exercise

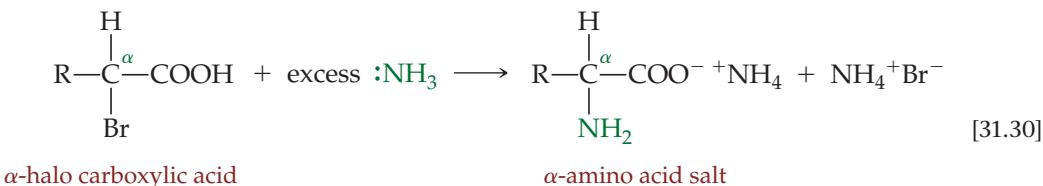
Estimate the net charge on serine at pH = 7 and pH = 10.

Reactions Involving Amino Acids

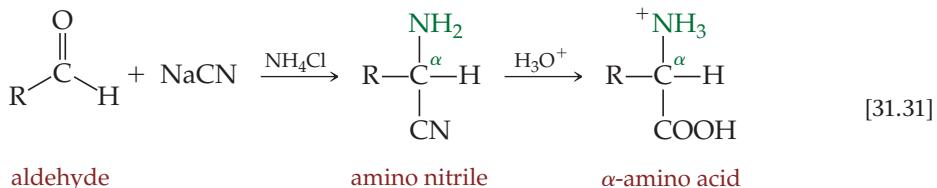
There are several biological pathways for synthesising amino acids. For example, the bacterium *Bacillus subtilis* synthesizes alanine from pyruvic acid in the following way:



In a laboratory, there are two main ways of synthesising α -amino acids. The first method involves nucleophilic substitution of an α -halo carboxylic acid with ammonia:



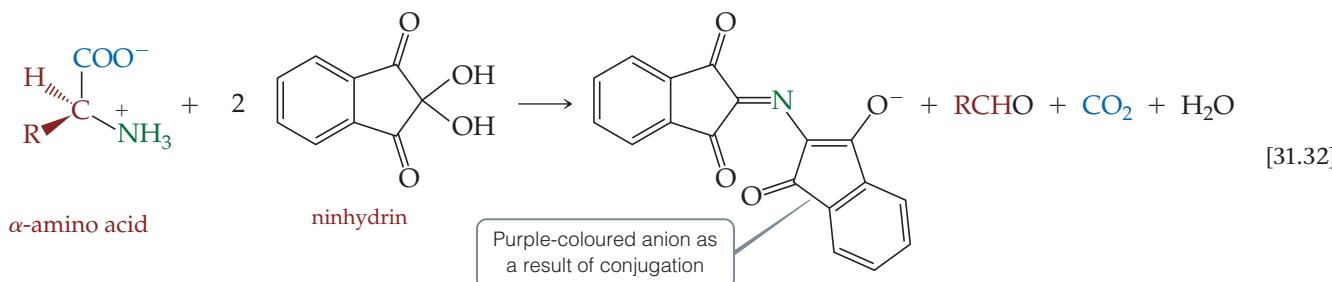
The second method is called **Strecker synthesis** of α -amino acids. This three-step reaction converts aldehydes to α -amino acids in the presence of cyanide ion.



Treatment of the aldehyde with cyanide ion in the presence of ammonium chloride yields the intermediate amino nitrile. This reaction is twofold: reaction of the aldehyde with ammonium chloride yields an imine, which undergoes addition of the cyanide ion to form the amino nitrile; the nitrile group is then hydrolyzed to the carboxylic acid in the presence of aqueous acid.

The methods described in Equations 31.30 and 31.31 are not ideal because the products are typically formed as racemates, which need to be separated to form the enantiomerically pure compounds. In some instances, separation is easily achieved by chiral resolution techniques. In other cases, more complicated and expensive techniques are required.

Since most amino acids are neither colored nor fluorescent, their detection requires the use of a dye. The most common dye used, called ninhydrin, reacts specifically with α -amino acids to produce a purple-colored anion, an aldehyde, CO_2 and water.



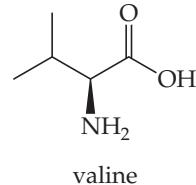
The fact that a purple-colored solution is formed from colorless starting materials is important for the use of ninhydrin in quantifying the amount of α -amino acid present in a sample using absorption spectroscopy at 580 nm. Note that only α -amino acids containing a primary amino group react to form the purple-colored anion. Proline, which has a secondary amino group, does not react in the same way.

Self-Assessment Exercises

- 31.12** Methionine has pK_a values of 2.28 and 9.21. At what pH do you get the maximum concentration of the zwitterionic form of methionine?

- (a) 2.28
- (b) 5.75
- (c) 7.00
- (d) 9.21

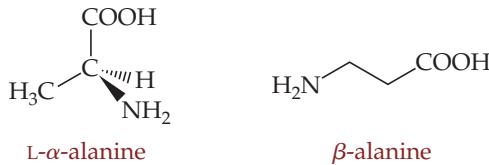
- 31.13** In the Strecker synthesis of valine, what is the starting aldehyde?



- (a) 3-methylbutanoic acid
- (b) 3-methylbutanal
- (c) methylpropanal

Exercises

- 31.14** Describe the relationship between β -alanine and L- α -alanine.



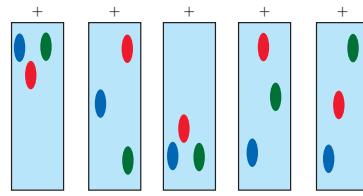
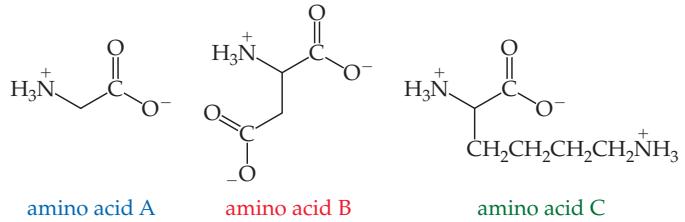
- 31.15** Alanine can exist in water in several ionic forms. (a) Suggest the form of alanine at low pH and at high pH. (b) Amino acids are reported to have two pK_a values, one in the range of 2 to 3 and the other in the range of 9 to 10. Alanine, for example, has pK_a values of about 2.3 and 9.6. Suggest the origin of the two pK_a values.

- 31.16** Glutamic acid has pK_a values of 2.19, 4.25 and 9.67. (a) What would be the predominant species in solution at pH 2? (b) What would be the predominant species in solution at pH 8? (c) Estimate the isoelectric point, pI , for glutamic acid.

- 31.17** Using Equation 31.28, estimate the average charge on glutamine (a) at pH 3, (b) at pH 7. Glutamine has pK_a values of 2.17 and 9.13.

- 31.18** Why is the average pK_a value (pK_{a1}) for the carboxyl group of an α -amino acid ~ 2 when that for acetic acid is about 4.8?

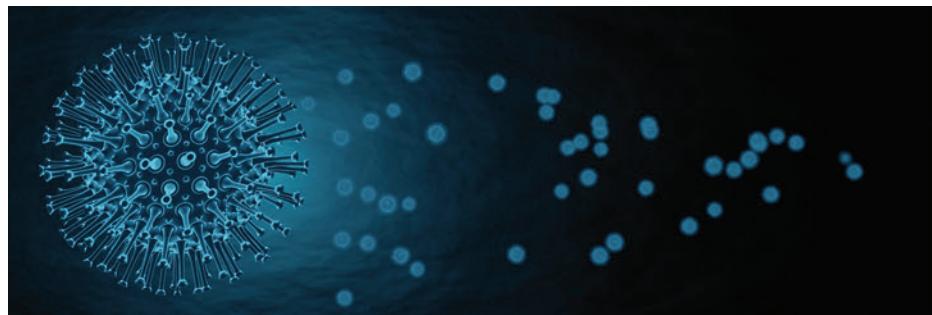
- 31.19** Which of the following electrophoresis results could represent the separation of the following three compounds if a mixture is applied to the center of the strip? Explain.



- 31.20** Outline a synthesis of phenylalanine (a) from 2-bromo-3-phenylpropanoic acid and (b) by the Strecker synthesis from 2-phenylethanal.

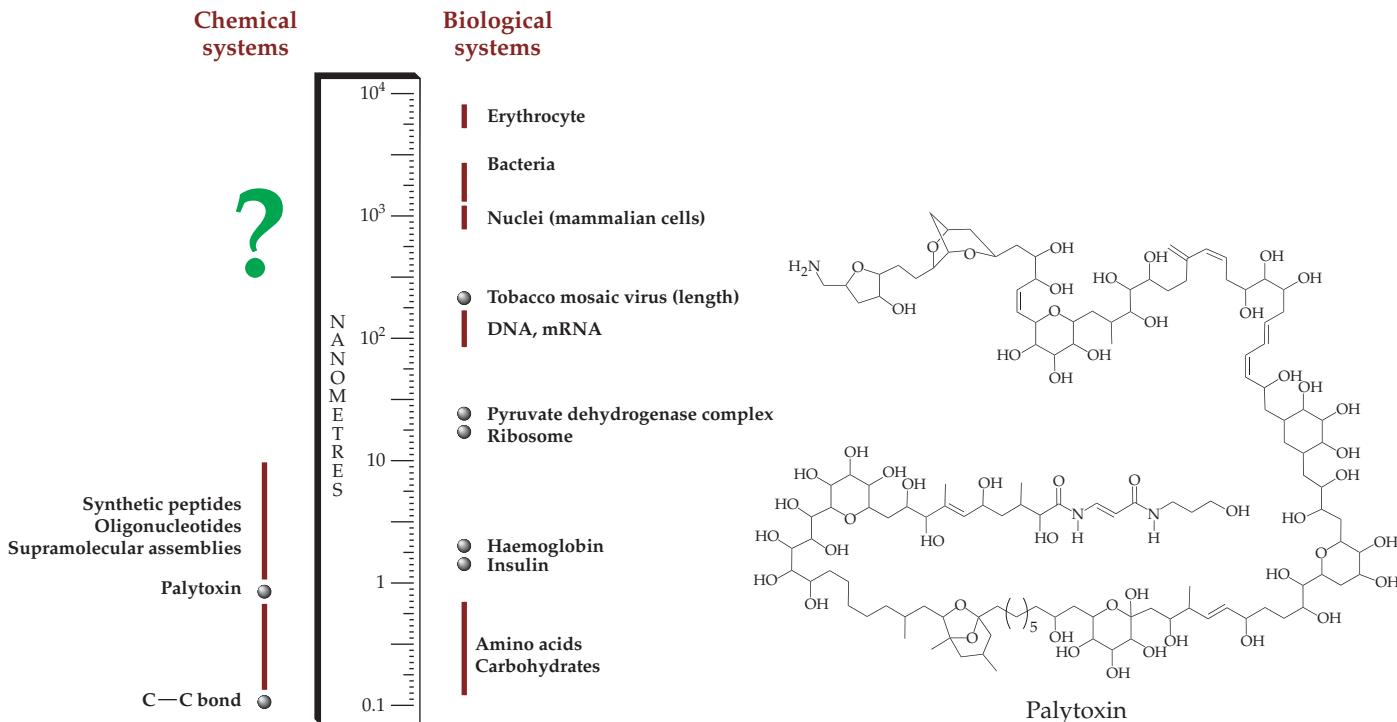


31.3 | Proteins, Peptides, and Enzymes



Research at the interface of chemistry and biology, termed *biological chemistry*, is an exciting and dynamic area for scientists in both disciplines. Biological chemists apply the principles of chemistry to understand the molecular basis of biological processes relevant to medicine, biotechnology, and other life sciences, such as structural biology, biochemistry, genetics, pharmacology, and physiology. They then apply this knowledge to new chemical transformations, in materials science, or in the area of medicinal chemistry. This differs from the area of science known as *biochemistry*, which aims to understand life in terms of its processes and reactions, but not necessarily on a molecular scale. In the next two sections, we present a brief overview of some of the elementary aspects of biological chemistry—proteins, vitamins, and nucleic acids—from a molecular point of view and discuss the importance of nitrogen in each case.

Many biologically important molecules are very large. Nevertheless, they may be understood by first breaking them down into their component parts, each containing the functional groups that we have studied in the last five chapters. The synthesis of large molecules is remarkable and one that places large demands on the chemical systems in



living organisms. Organisms build complex biomolecules (as, for example, in the coronavirus disease COVID-19) from much smaller and simpler substances that are readily available in the biosphere. The synthesis of large molecules requires energy because most of the reactions required are endothermic and also represent a decrease in entropy as a random mix of raw materials is assembled into an ordered structure. The ultimate source of this energy is the sun. Mammals and other animals have no capacity for using solar energy directly; rather, they depend on plant photosynthesis to supply the bulk of their energy needs.

By the end of this section, you should be able to:

- Understand the components of protein structure
- Know how a protein sequence may be determined

We have introduced some important biochemical applications of fundamental chemical ideas in the Chemistry and Life boxes that appear throughout this text. Hydrogen bonding, for example, is critical to the function of many biological systems. The geometry of, and functional groups within, molecules can govern their biological activity. Many of the large molecules in living systems are polymers of much smaller molecules. These polymers, called *biopolymers*, can be classified into three broad categories: proteins, polysaccharides (carbohydrates) and nucleic acids.

Proteins are macromolecular substances present in all living cells and viruses. About 50% of your body's dry weight is protein. Proteins serve as the major structural components in animal tissues; they are a key part of skin, nails, cartilage and muscles. Other proteins catalyze reactions, transport oxygen, serve as hormones to regulate specific body processes and perform other tasks. Whatever their function, all proteins are chemically similar, being composed of the same basic building blocks—amino acids.

Proteins can be classified in two ways, as simple or conjugated proteins or as globular and fibrous proteins. **Simple proteins** are those that yield only amino acids upon hydrolysis; for example, blood serum albumin. **Conjugated proteins** yield other compounds, such as carbohydrates, lipids or metal complexes, as well as amino acids on hydrolysis. Examples include myoglobin and ferrichrome. In fact, conjugated proteins are more common in the body than simple proteins. **Globular proteins** are tightly bundled proteins that are named because of their globular or spherical shape. These proteins, which include albumins, enzymes, immunoglobulins and insulin, are water soluble and typically mobile within cells. They have non-structural functions, such as combating the invasion of foreign objects, transporting and storing oxygen and acting as catalysts. **Fibrous proteins**, such as collagen and keratins, are arranged like filaments or fibers. In these substances the long coils align themselves in a more or less parallel fashion to form long, water insoluble fibers. Fibrous proteins provide structural integrity and strength to many kinds of tissue and are the main components of muscle, tendons and hair. **Table 31.1** summarizes some common protein classes and their functions.

Peptide is the name given to a short polymer (oligomer) of amino acids joined by *peptide amide bonds*. Peptides are classified by the number of amino acids in the chain. For example, a *dipeptide* is a molecule containing two amino acids joined by a peptide bond. A *tripeptide* is a molecule containing three amino acids joined by peptide bonds.

Polypeptides are formed when many amino acids are linked together by peptide bonds. They are classed as *macromolecules* or *biopolymers*.

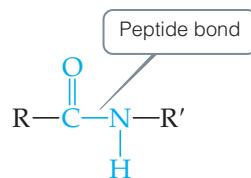
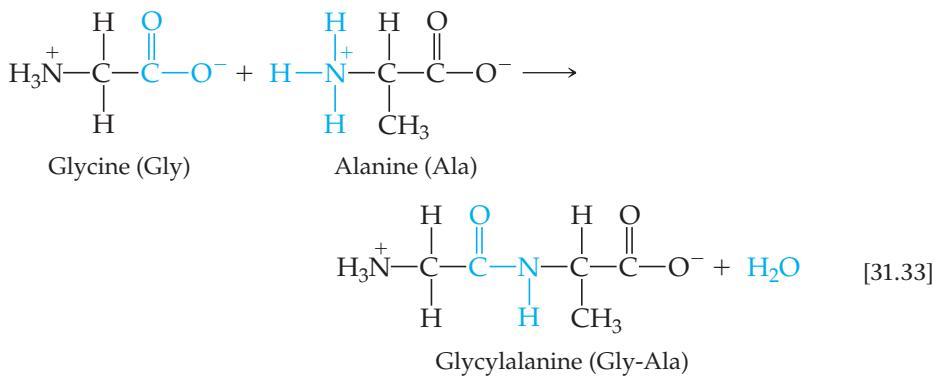


TABLE 31.1 Some common protein classes and their function

Name of class	Composition examples	Definition and functions
Glycoproteins	Contain carbohydrates, e.g. ovalbumin (egg white)	A ubiquitous family of proteins containing oligosaccharides (carbohydrates), which impart properties of solubility, ligand affinity, cellular targeting and stability. Many proteins released by cells to the blood and other fluids are glycoproteins. For example, Antarctic fish survive near-freezing water temperatures as a result of freezing-point depression of their blood serum by a globular glycoprotein.
Lipoproteins	Contain fats, oils or steroids (lipids), e.g. low-density lipoproteins (LDLs) and high-density lipoproteins (HDLs)	Lipoproteins serve as carriers of hydrophobic ligands through solubility barriers such as aqueous body fluids. LDLs are the main transport for cholesterol through the body. HDLs carry excess cholesterol to the liver for processing.
Metalloproteins	Contain coordinated metal ions, e.g. hemoglobin, ferredoxin, NADH dehydrogenase, coenzyme Q, cytochrome c reductase	Metalloproteins are proteins that contain a metal cofactor. The metal may be an isolated ion or may be coordinated with a non-proteinaceous organic compound, such as the porphyrin found in hemoproteins. In some cases, the metal is coordinated with a side chain of the protein and an inorganic non-metallic ion.
Nucleoproteins	Contain nucleic acids, e.g. chromatin, RNA binding proteins, ribosomes	Nucleoproteins are any supramolecular complex of protein and nucleic acid; or any protein usually found closely associated with nucleic acid. They contain a diverse group of viral and genetic regulatory proteins which bind in a sequence-specific manner to DNA.
Phosphoproteins	Contain a phosphate group, e.g. synapsins, caesins (milk)	Phosphoproteins are any of a group of proteins containing chemically bound phosphoric acid. Protein phosphorylation is probably the most ubiquitous and diverse molecular mechanism of biological signaling and regulation.

Coding Peptides

A **peptide bond** is formed by a condensation reaction between the carboxyl group of one amino acid and the amino group of another amino acid. Alanine and glycine, for example, can react to form the dipeptide glycylalanine:



The amino acid that furnishes the carboxyl group for peptide bond formation is named first, with a *-yl* ending; then the amino acid furnishing the amino group is named second. To make the naming of peptides easier, especially those that are more than 25 amino acids long, series of three-letter and single-letter codes have been devised (Table 31.2). Based on the three-letter codes for the amino acids from Table 31.2, glycylalanine can be abbreviated Gly-Ala. In this notation it is understood that the unreacted amino group is on the left and the unreacted carboxyl group on the right. The artificial sweetener *aspartame* (Figure 31.19) is the methyl ester of the dipeptide of aspartic acid and phenylalanine, Asp-Phe:

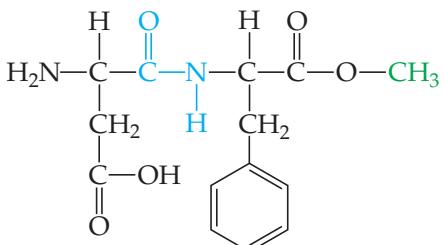


TABLE 31.2 Amino acid codes for proteins and peptides

Amino acid	Three-letter code	Single-letter code
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Leucine	Leu	L
Isoleucine	Ile	I
Methionine	Met	M
Phenylalanine	Phe	F
Tryptophan	Trp	W
Proline	Pro	P
Asparagine	Asn	N
Glutamine	Gln	Q
Serine	Ser	S
Threonine	Thr	T
Aspartic acid	Asp	D
Glutamic acid	Glu	E
Cysteine	Cys	C
Tyrosine	Tyr	Y
Arginine	Arg	R
Histidine	His	H
Lysine	Lys	K

Proteins are linear (that is, unbranched) polypeptide molecules with molecular weights ranging from about 5000 to over 50 million u. Proteins differ from polypeptides in that proteins are capable of some function, whether chemical or structural, as a result of their three-dimensional structure. Because 20 different amino acids are commonly linked together in proteins and because proteins consist of hundreds of amino acids, the number of possible arrangements of amino acids within proteins is virtually limitless.



Sample Exercise 31.7

Drawing the Structural Formula of a Tripeptide



Draw the full structural formula for alanylglycylserine.

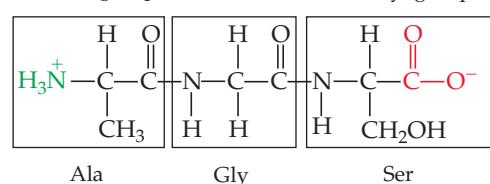
SOLUTION

Analyze We are asked to draw a tripeptide.

Plan The name of this substance suggests that three amino acids—alanine, glycine and serine—have been linked together, forming a *tripeptide*. Note that the ending *-yl* has been added to each amino acid except for the last one, serine. By convention, the first-named amino acid (alanine, in this case) has a free amino group and the lastnamed one (serine) has a free carboxyl group. Thus, we can construct the structural formula of the tripeptide from its amino acid building blocks.

Solve We first combine the carboxyl group of alanine with the amino group of glycine to form a peptide bond and then the carboxyl group of glycine with the amino group of serine to form another peptide bond. The resulting tripeptide consists of three “building blocks” connected by peptide bonds:

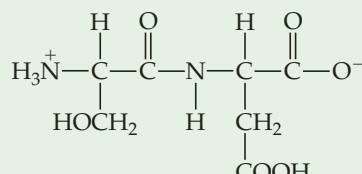
Amino group → Carboxyl group



We can abbreviate this tripeptide as Ala-Gly-Ser, or AGS.

► Practice Exercise

Name the dipeptide that has the following structure and give its abbreviation:

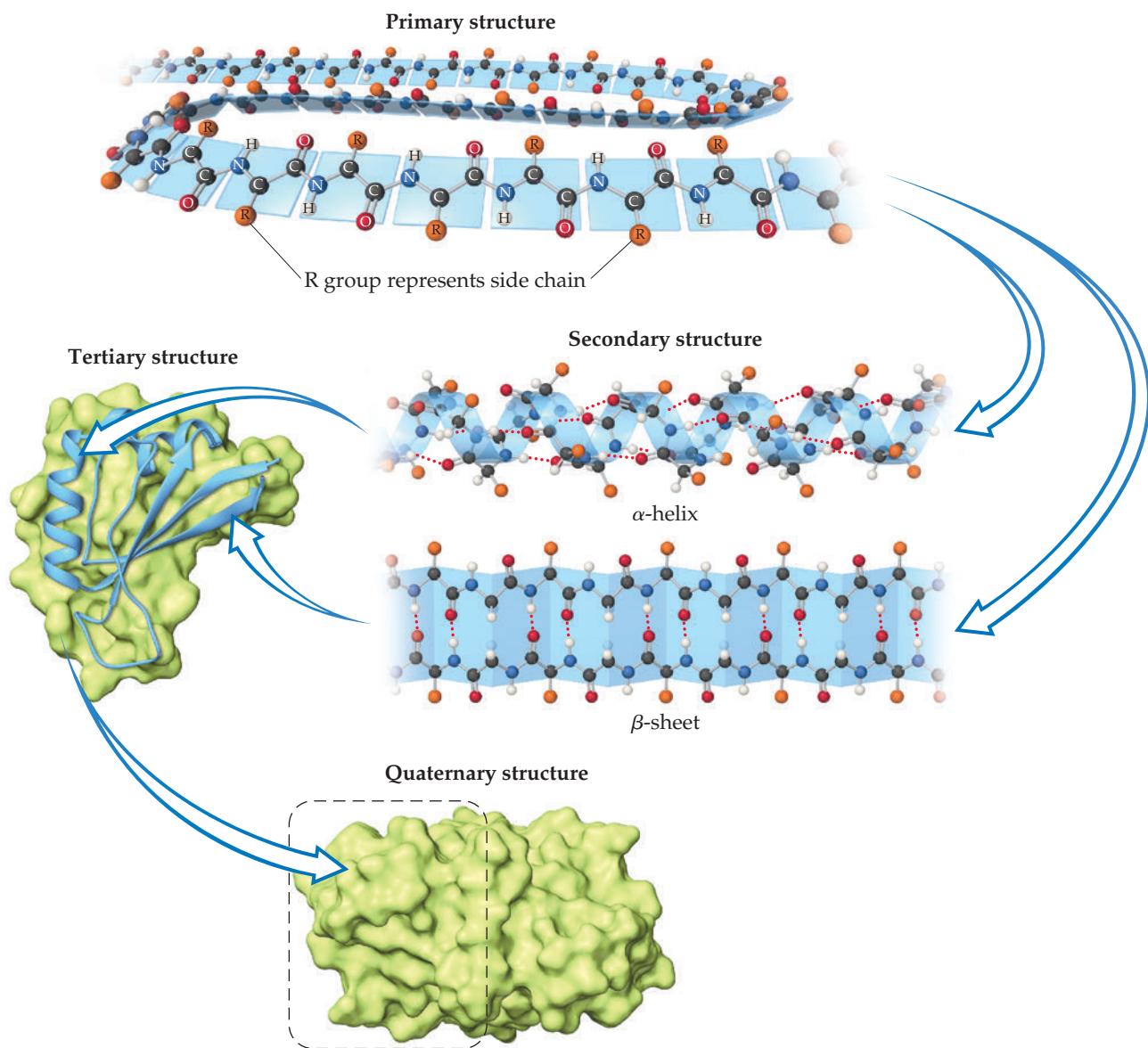


Protein Structure

The arrangement, or sequence, of amino acids along a protein chain is called its **primary structure**. The primary structure gives the protein its unique identity. A change in even one amino acid can alter the biochemical and physiological characteristics of the protein. For example, sickle-cell anemia is a genetic disorder resulting from a single replacement in a protein chain in hemoglobin. The chain that is affected contains 146 amino acids. The substitution of a single amino acid with an acidic functional group for one that has a hydrocarbon side chain alters the solubility properties of the hemoglobin, and normal blood flow is impeded (see the A Closer Look box in Section 31.2).

Another example occurs within the free-radical scavenging Zn/Cu enzyme, superoxide dismutase (SOD1). A single mutation in the primary structure of SOD1—for example, by replacing Gly with Ala at the 93rd position of the peptide sequence—is enough to implicate this mutant enzyme as a contributor to motor neuron disease (MND).

Proteins in living organisms are not simply long, flexible chains with random shapes. Rather, the chains coil and stretch in particular ways. The **secondary structure** of a protein refers to how segments of the protein chain are orientated in a regular pattern as seen in [Figure 31.20](#).



▲ **Figure 31.20** The four levels of structure of proteins.

One of the most important and common secondary-structure arrangements is the **α -helix** (Figure 31.20), first proposed by Linus Pauling and Robert B. Corey (1897–1971). Imagine winding a long protein chain in a helical fashion around a long cylinder. The helix is held in position by hydrogen bond interactions between N—H bonds and the oxygen atoms of nearby carbonyl groups. The pitch of the helix and the diameter of the cylinder must be such that (1) no bond angles are strained and (2) the N—H and C=O functional groups on adjacent turns are in a proper position for hydrogen bonding. In fact, each C=O group is hydrogen bonded to an N—H group four amino acid units away from it. An arrangement of this kind is possible for some amino acids along the chain, but not for others. Large protein molecules may contain segments of the chain that have the α -helical arrangement, interspersed with sections in which the chain is in a random coil. Typically, the α -helix has 3.6 amino acids per turn. The N—H groups of the amino acid point along the axis of the helix and all orientate themselves in the same direction. The C=O groups of the amino acids also point along the axis but in the opposite direction to the N—H groups, as a result of the *trans* arrangement of the peptide bond. All R groups point away from the helix structure.

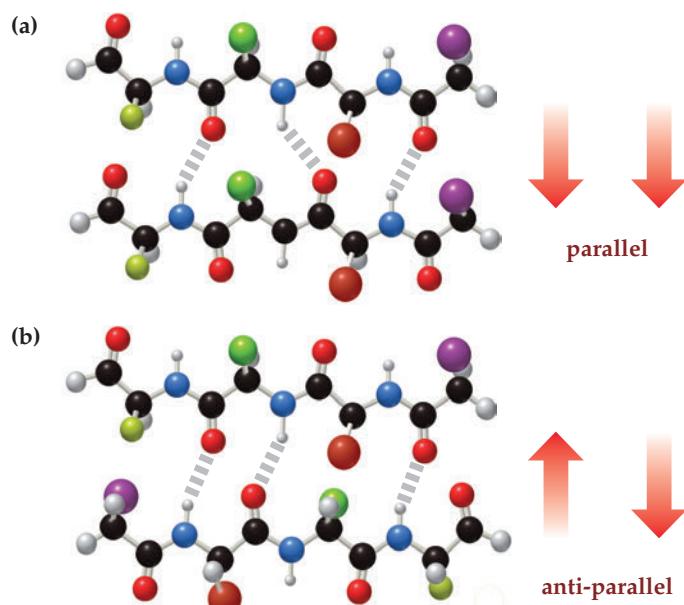
Figure 31.21 shows the structure of an enzyme containing a small molecular substrate. The important aspect of this protein is the highly coiled nature of the amino acid backbone (depicted as a ribbon). These coils are α -helices, which help give rise to the overall three-dimensional structure of the protein.

Another form of secondary structure that is often seen in proteins is the **β -pleated sheet**. This structure is not restricted to *intramolecular* hydrogen bonding structure, as is the α -helix, but may also form an *intermolecular* hydrogen bonding arrangement with another peptide chain (Figure 31.22). Since the peptide backbone contains only the α -carbon, NH and carbonyl carbon in an alternating manner from the N- to C-terminus, the two peptide chains can hydrogen bond in one of two orientations: parallel or anti-parallel. Both orientations are seen in natural proteins, although anti-parallel is the most common form. Figure 31.23 shows an example of a protein with extensive β -sheets. These β -sheets are formed by the interaction of two or more peptide chains in an anti-parallel manner (as indicated by the direction of the arrows). Note that it is not necessary for all the peptide chains contributing to the sheet structure to be of the same size.

Proteins are not active biologically unless they are in a particular form in solution. The process by which the protein adopts the biologically active form is called **folding**.



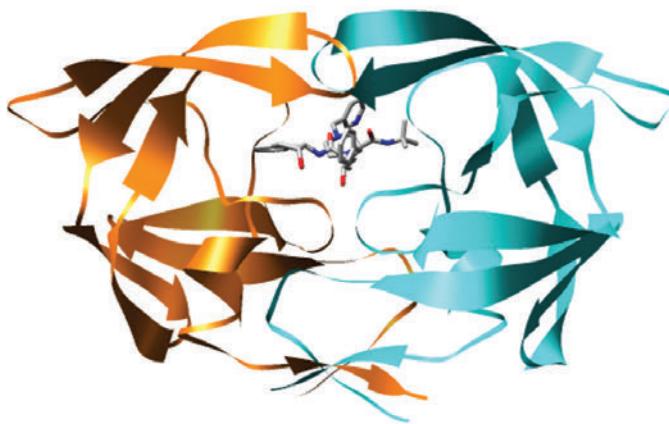
▲ **Figure 31.21** An enzyme containing a substrate. A computer-generated model of an enzyme showing the amino acid backbone as a green ribbon. The substrate (violet) is shown to interact with a specific part of the protein called the active site.



◀ **Figure 31.22** β -sheet form of a protein. F, Cl, Br and I are used along the peptide backbone to describe the two orientations possible for a β -sheet formation. (a) A parallel sheet, in which the two peptides orientate in the same direction from N- to C-terminus. (b) An anti-parallel sheet in which the two halide trends (F to I) occur in opposite directions.

► **Figure 31.23 A dimeric protein containing β -sheets.**

The arrowed ribbons represent the directions of peptide chains that undergo intermolecular hydrogen bonding, forming β -sheet structures. The substrate in the middle fits nicely into the void made by the two halves of this protein dimer.



The overall shape of a protein in its folded form, determined by all the bends, kinks and sections of rod-like α -helical and flat β -sheet structures, is called the **tertiary structure**.

The tertiary structure of a protein is maintained by many different interactions. Certain folding of the protein chain leads to lower-energy (more stable) arrangements than other folding patterns. For example, a globular protein dissolved in aqueous solution folds in such a way that the non-polar hydrocarbon portions are tucked within the molecule, away from the polar water molecules. Most of the more polar acidic and basic side chains, however, project into the solution, where they can interact with water molecules through ion-dipole, dipole-dipole or hydrogen bonding interactions. The result of a mis-folding based on the change of a polar group was seen clearly in the example of sickle-cell anemia.

Some proteins are more complicated and have a fourth level of structure known as the **quaternary structure**. This level of structure involves the aggregation of two or more protein substructures into a larger macromolecular assembly. For example, the protein shown in Figure 31.23 has a quaternary structure, made up of two identical protein substructures that interact non-covalently with each other. The function the protein exhibits is only possible due to the quaternary (dimeric) structure and not by either substructure separately.

Enzymes

One of the most important classes of proteins are the **enzymes**, large protein molecules that serve as catalysts. Enzymes usually catalyze only very specific reactions. Their tertiary and quaternary structures generally dictate that only certain substrate molecules can interact with the enzyme (Figures 31.21 and 31.23).

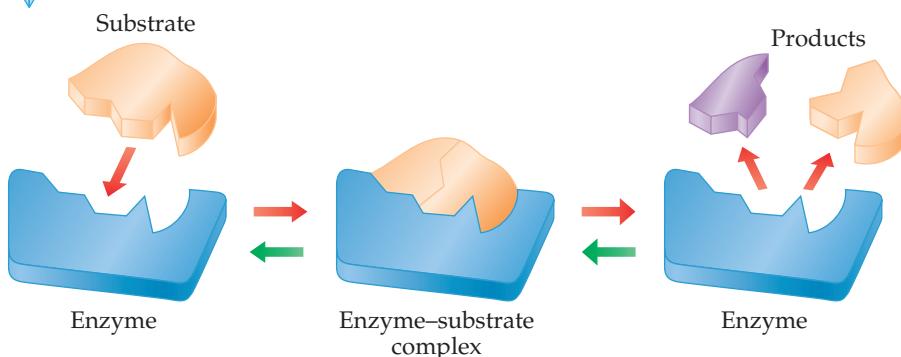
Many of the most interesting and important examples of catalysis involve reactions within living systems. The human body is characterized by an extremely complex system of interrelated chemical reactions. All these reactions must occur at carefully controlled rates and at 37 °C in order to maintain life. Enzymes enable these reactions to take place in the proper order and at the correct rate. Most enzymes are proteins with molecular weights ranging from about 10 000 to about 1 million u. They are very selective in the reactions they catalyze, and some are absolutely specific, operating for only one substance in only one reaction. The decomposition of hydrogen peroxide, for example, is an important biological process. Because hydrogen peroxide is strongly oxidizing, it can be physiologically harmful. For this reason, the blood and livers of mammals contain an enzyme, *catalase*, which catalyzes the decomposition of hydrogen peroxide into water and oxygen (Equation 15.33). Figure 31.24 shows the dramatic acceleration of this chemical reaction by the catalase in beef liver.



▲ **Figure 31.24 Effect of an enzyme.** Ground-up beef liver causes hydrogen peroxide to decompose rapidly into water and oxygen. The decomposition is catalyzed by the enzyme *catalase*. Grinding the liver breaks open the cells, so that the reaction takes place more rapidly. The frothing is due to escape of oxygen gas from the reaction mixture.

Go Figure

Why might this be called a lock-and-key model?

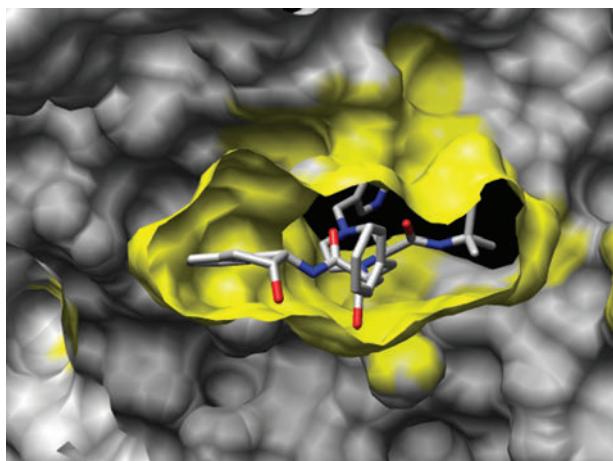


▲ Figure 31.25 The lock-and-key model for enzyme action. The correct substrate is recognized by its ability to fit the active site of the enzyme, forming the enzyme–substrate complex. After the reaction of the substrate is complete, the products separate from the enzyme. The reaction is just as suited to forming covalent bonds (right to left) as it is in breaking them (left to right).

Although an enzyme is a large molecule, the reaction is catalyzed at a very specific location in the enzyme, called the **active site**. The substances that undergo reaction at this site are called **substrates**. A simple explanation for the specificity demonstrated by enzymes is provided by the **lock-and-key model** illustrated in **Figure 31.25**. The substrate is pictured as fitting neatly into a special place on the enzyme (the active site), much like a specific key fitting into a lock. The active site is created by the coiling and folding of long protein chains to form a space, something like a pocket, into which the substrate molecule fits. **Figure 31.26** shows a model of an enzyme’s active site (cross-sectioned) with a bound substrate molecule.

The combination of the enzyme and the substrate is called the enzyme–substrate complex. Although Figure 31.25 shows both the active site and its complementary substrate as having rigid shapes, there is often a fair amount of flexibility in the active site. Thus, the active site may change shape as it binds the substrate. The binding between the substrate and the active site involves intermolecular forces such as dipole–dipole attractions, hydrogen bonds and London dispersion forces.

As the substrate molecules enter the active site, they are activated so that they are capable of extremely rapid reaction at mild (for example, 37 °C) temperatures. This activation may result from the enzyme withdrawing or donating electron density at a particular bond. In addition, in the process of fitting into the active site, the substrate molecule may be distorted and thus made more reactive. Once the reaction occurs, the products then depart, allowing another substrate molecule to enter. This means that the products must



▲ Figure 31.26 Molecular model of an active site. The active site (shown in yellow) has been cut away to show how well the substrate fits into the different depressions and pockets, like a hand in a glove. Compare this representation to that in Figure 31.23 of the same protein.

bind more weakly to the active site than does the substrate. If the reverse were true, inhibition of the enzyme's activity would be observed.

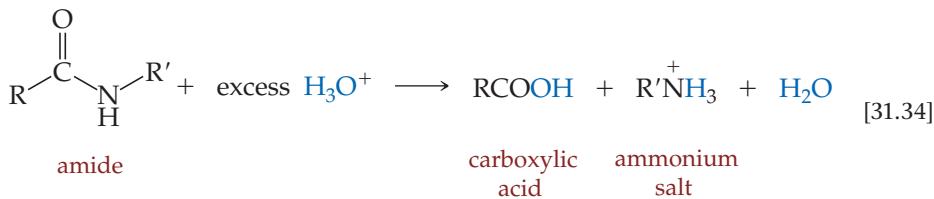
The activity of an enzyme is also destroyed if some molecule in the solution is able to bind strongly to the active site and block the entry of the substrate. Such substances are called *enzyme inhibitors*. Nerve poisons and some toxic metal ions, such as lead and mercury, are believed to act in this way to inhibit enzyme activity. Other poisons act by attaching elsewhere on the enzyme, thereby distorting the active site so that the substrate no longer fits.

Enzymes are typically more efficient than human-made catalysts. The number of individual catalyzed reaction events occurring at a particular active site, called the *turnover number*, is generally in the range of 10^3 to 10^7 per second. Such large turnover numbers correspond to very low activation energies. This is an area of chemistry of intense study because of its benefits associated with (a) mild reaction conditions, (b) high turnover numbers, (c) the sustainability of using microbes, bacteria and other organisms to promote reaction and (d) the removal of hazardous chemicals and solvents. Remember, most biological processes occur in water.

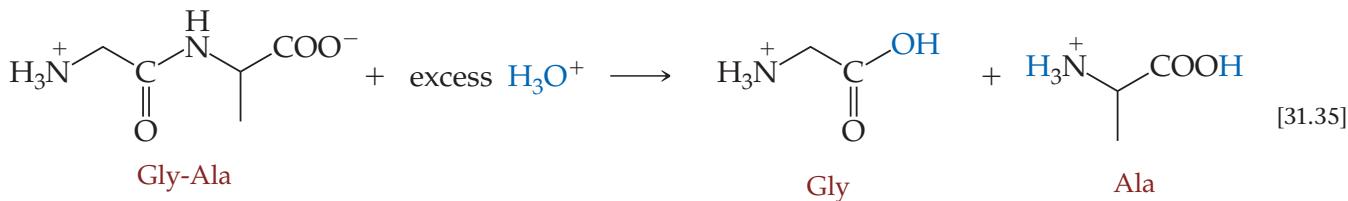
Sequencing of Peptides and Proteins

As a result of the almost endless permutations of amino acid sequences and the very specific nature of a protein's function, chemists and biochemists have come up with some clever ways of determining the precise structure of a protein. Let's consider the problem. There are three aspects of sequencing that need to be elucidated: (1) which amino acids are present; (2) what amounts of each amino acid are present; and (3) what order do these amino acids take in the primary structure of the protein.

The first two of these questions are easily answered provided the protein is pure. Amide bonds are hydrolyzed in excess aqueous 6M HCl to yield the ammonium salt and carboxylic acid:

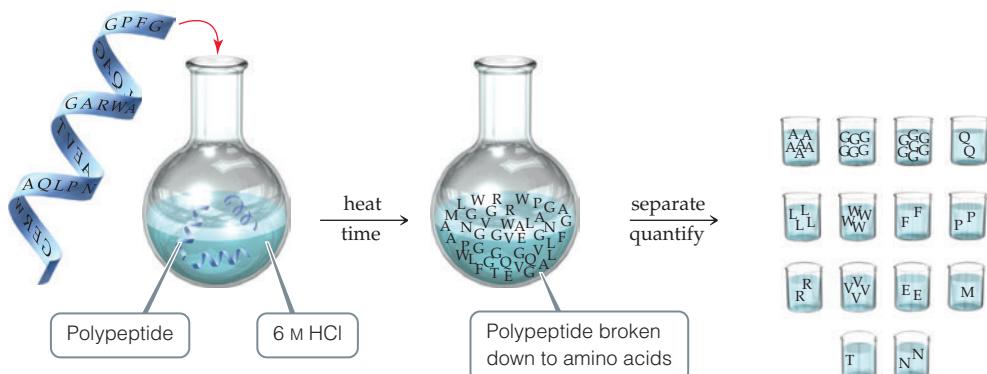


For a simple dipeptide such as Gly-Ala, the products are the ammonium salts of glycine and alanine:



Amino acids can be separated from each other by electrophoresis (Figure 31.18) and other separating methods such as chromatography. The amount of each amino acid is then quantified using ninhydrin by the reaction shown in Equation 31.32. **Figure 31.27** summarizes this entire process using a general protein.

Knowing which amino acids are present, as well as their relative amounts, is only the beginning. The specific sequence of amino acids within the protein, which gives rise to the protein's function, is critical. Even, the simple dipeptide Gly-Ala is different from Ala-Gly. The processes required to sequence a protein need to be very specific and efficient. Two approaches are possible. Either the whole protein can be sequenced, or the protein can be cleaved at specific points to make smaller peptides, which are then sequenced. Both approaches have their advantages, although the latter is more practical. **Table 31.3** lists three common methods of undertaking the cleavage of peptide bonds at specific amino acid sites.

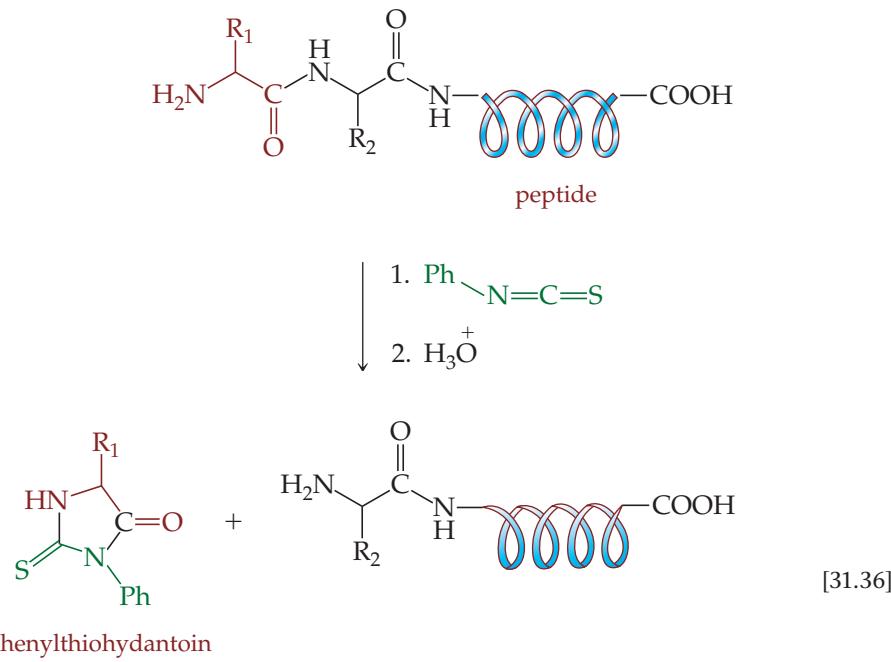
**◀ Figure 31.27** Amino acid analysis.

Note that the single-letter codes for the amino acids (Table 31.2) have been used in this example.

TABLE 31.3 Useful reagents for protein cleavage

Reagent	Cleavage point(s) at the carboxyl group of
Cyanogen bromide (BrCN)	Methionine
Trypsin	Arginine and lysine
Chymotrypsin	Phenylalanine, tryptophan and tyrosine
6 M HCl (partial hydrolysis)	Indiscriminant

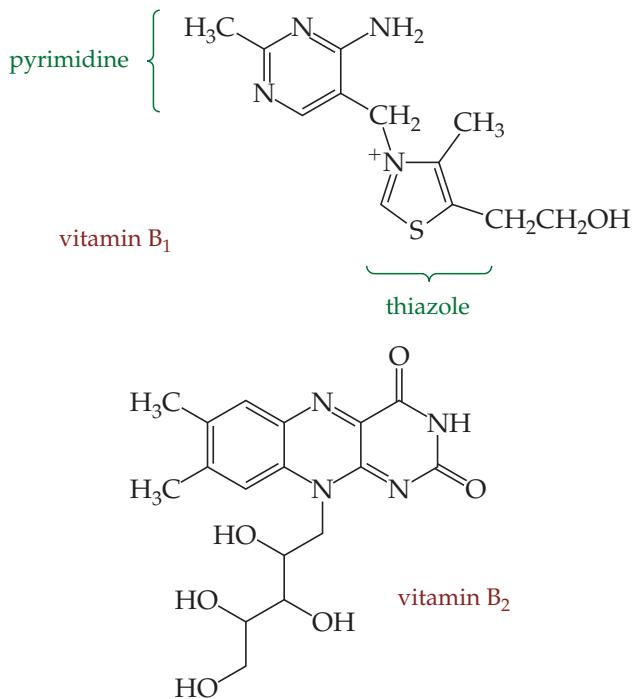
The most widely used method of sequencing proteins was developed by Pehr Edman (1916–1977) in the 1950s. The process is a laborious one and has been automated to save time and sample amounts. Modern instruments can sequence as little as $0.1\ \mu\text{g}$ of a protein. The general idea of his approach, called the **Edman degradation**, is to cleave and identify one amino acid at a time from the *N*-terminus of the peptide chain. The process is repeated until the entire sequence is known. The Edman degradation requires treatment of the peptide with phenyl isothiocyanate ($\text{Ph}-\text{N}=\text{C}=\text{S}$), followed by mild acid hydrolysis. The product of this reaction, a phenylthiohydantoin, is collected and compared with known phenylthiohydantoin derivatives of the 20 common amino acids to identify which amino acid is cleaved. The new peptide (which is now one residue shorter) is also subjected to the same conditions to characterize the second amino acid residue and so on.



CHEMISTRY AND LIFE B Group Vitamins

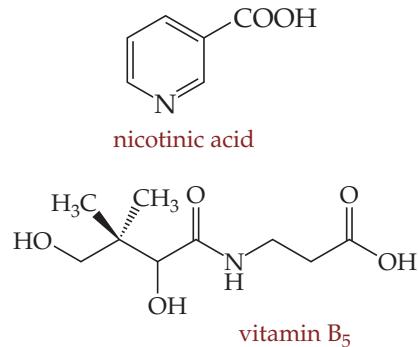
Vitamins are small organic molecules that must be obtained through diet. They are usually required in trace amounts for proper growth. Many vitamins are used as coenzymes by the body; that is, they are required for the catalytic function of certain enzymes. All of the major B-group vitamins are nitrogen-containing compounds.

Vitamin B₁ is also known as thiamin, a water soluble heterocyclic structure made up of a substituted pyrimidine and a thiazole ring coupled by a methylene (CH₂) bridge. Thiamin is converted to its active form, thiamin pyrophosphate (TPP), in the brain and liver. In this form, thiamin acts as a cofactor for reductase enzymes responsible for, among other things, the reduction of pyruvate (Equation 31.29). A deficiency in thiamin may lead to *fatigue* and *depression*. Riboflavin, or vitamin B₂, is employed as a coenzyme for flavoproteins, which have a wide range of redox roles within the body. Riboflavin is found in useful quantities in eggs, milk, meat and cereal. A deficiency in vitamin B₂ leads, among other things, to *cracked lips* and *scaly skin*.

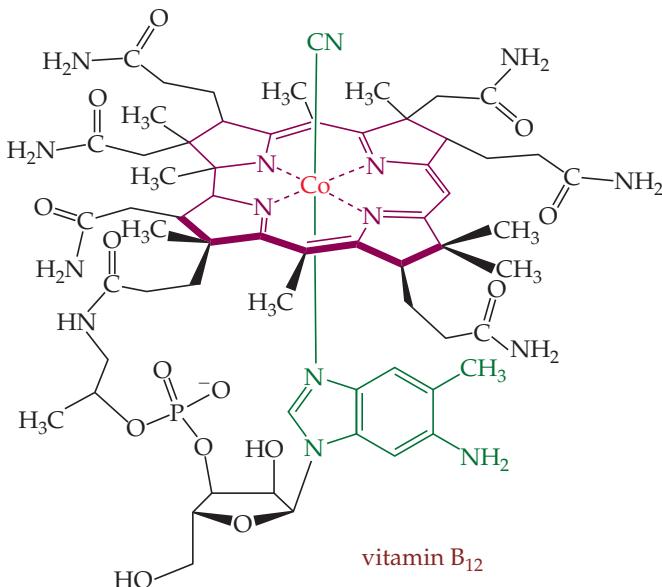


Nicotinic acid is the most common source of Vitamin B₃, also known as niacin. The human body uses vitamin B₃ in the form of NAD⁺ (nicotinamide adenine dinucleotide) and NADP⁺ (nicotinamide adenine dinucleotide phosphate) as a coenzyme in more than 50 enzyme-based reactions. Vitamin B₃ is instrumental in the release of energy from carbohydrates. It is necessary for proper brain function and is also involved in fat and cholesterol metabolism. A deficiency in vitamin B₃ may lead to *dermatitis*, *diarrhoea* and *dementia*. Vitamin B₅, also known as pantothenic acid, is found in abundance in whole-grain cereals, legumes and meat, and is usually formed from β -alanine and pantoic acid. Vitamin B₅ is required for the synthesis of coenzyme A

(CoA), an acyltransferase, and for the metabolism of carbohydrates, fats and proteins. A deficiency in vitamin B₅ leads to *weight loss* and *irritability*.



One of the most complicated and essential of the vitamins is vitamin B₁₂. Vitamin B₁₂ is unusual as it is the only metal-ion-containing vitamin. The tetrapyrrole macrocycle (purple) of vitamin B₁₂ is called a *corrin*, which is related to the porphyrin structure found in myoglobin, hemoglobin and chlorophyll. Centrally coordinated to the corrin macrocycle is a cobalt(II) ion with an octahedral geometry (red). The cyano (CN) and benzimidazole ligands (green) bind axially to the metal ion.



Vitamin B₁₂ is synthesized exclusively by microorganisms and is found in the liver of animals. The liver can store up to six years' supply of vitamin B₁₂, so deficiencies in this vitamin are rare. Vitamin B₁₂ helps to protect nerves and is involved in the formation of red blood cells in bone marrow which prevents anemia. Vitamin B₁₂ is necessary in the catabolism of fatty acids.



Sample Exercise 31.8

Sequencing Peptides

An unknown pentapeptide was subjected to a range of different experiments to ascertain its structure. The following results were obtained:

Experiment	Fragment result
Total hydrolysis	Arg, Glu, His, Phe, Ser
Edman degradation	Glu
Chymotrypsin hydrolysis	Glu-His-Phe, Arg-Ser
Trypsin hydrolysis	Glu-His-Phe-Arg, Ser

What is the pentapeptide sequence?

SOLUTION

Analyze We are asked to sequence a small peptide based on the fragments obtained by different chemical reactions.

Plan Identify the reactions and what type of fragment each leads to. Determine the number of different amino acids listed through the hydrolysis products. Construct the peptide one fragment at a time.

Solve Total hydrolysis yields five amino acids, which is the total required for the pentapeptide. Hence, each amino acid occurs only once in the sequence. Peptides are always written from the *N*- to *C*-terminus. The Edman degradation yields the *N*-terminus residue, Glu. Chymotrypsin hydrolyzes at the carboxyl end of

Phe, which indicates the first three residues are Glu-His-Phe. At this point, the sequence can be only one of two possibilities: Glu-His-Phe-Arg-Ser or Glu-His-Phe-Ser-Arg. Since trypsin cleaves at the carboxyl group of Arg, leaving the single amino acid residue, Ser, the pentapeptide sequence was:



► Practice Exercise

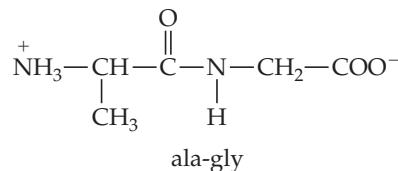
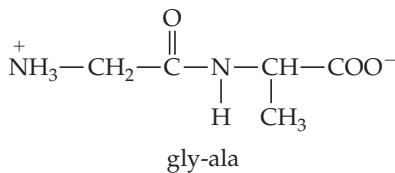
The partial hydrolysis of an unknown hexapeptide gave the following fragments: Gly-Ala-Phe, Phe-Tyr-His, Tyr-His-Glu. What is the sequence of the hexapeptide?

Self-Assessment Exercises

31.21 Does Gly-Ala have the same structure as Ala-Gly?

- (a) Yes
- (b) No

31.22 What information does the acid hydrolysis of a peptide provide on its composition or structure?



(a) The amino acids present

- (b) The primary structure of the peptide
- (c) The secondary structure of the peptide

Exercises

31.23 (a) Draw the two possible dipeptides formed by the condensation reaction between serine and lysine. (b) If a racemic mixture of both amino acids is used, how many stereoisomers will result?

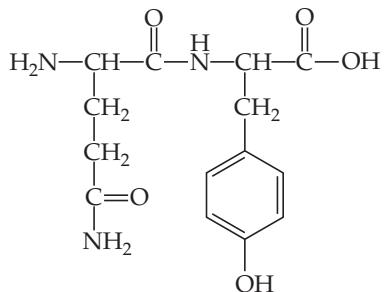
31.24 (a) Draw the condensed structure of the tripeptide Ile-Ala-Cys. (b) How many different tripeptides can be made from the amino acids serine, threonine and phenylalanine?

Give the abbreviations for each of these tripeptides, using the three-letter and one-letter codes for amino acids.

31.25 Describe what is meant by the primary, secondary and tertiary structures of proteins. Give examples to illustrate your answer.

31.26 Glutathione is a tripeptide found in most living cells. Partial hydrolysis yields Cys-Gly and Glu-Cys. What is the structure of glutathione?

- 31.27** Name the dipeptide that has the following structure, and give its one-letter and three-letter abbreviations:



- 31.28** Bradykinin is a small peptide containing nine amino acids. Partial hydrolysis gave the following fragments:

Arg-Pro-Pro, Ser-Pro, Pro-Phe-Arg, Gly-Phe-Ser, Pro-Pro-Gly

What is the primary structure of bradykinin?

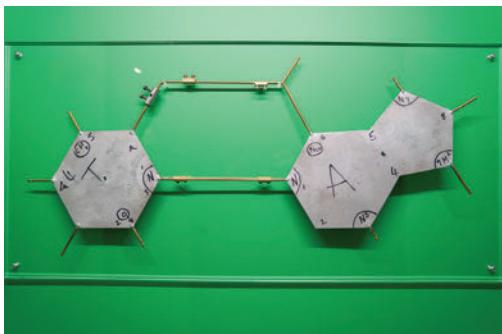
- 31.29** The abbreviation for a peptide is VAKG. **(a)** Draw the structural formula for this peptide. **(b)** Is this peptide neutral, acidic or basic? **(c)** What would be the products of a trypsin digestion? **(d)** Which amino acid would be identified in VAKG by the Edman degradation?

31.21 (b) 31.22 (a)

Answers to Self-Assessment Exercises



31.4 | Nucleic Acids and DNA



Gregor Mendel, who made a careful study of 8000 pea plants, was the first to recognize that there were some specific characteristics passed from one generation to the next. He published his findings in 1866, but it was not until forty years later that the term “gene” was coined. At this time, proteins were thought to be the repository of inherited information until the 1940’s when experiments concluded that DNA was fundamental to the process. There were many people contributing vital information that eventually led James Watson and Francis Crick to propose the structure of DNA in 1953. DNA was discovered in 1869 by Friedrich Miescher, while the Russian born biochemist Phoebus Levene, working in the early twentieth century in America, determined the component parts that made up this biopolymer—a sugar, a phosphate group, and four nitrogen heterocyclic compounds, referred to as “bases”. Erwin Chargaff studied these bases and found that, irrespective of the organism the DNA was extracted from, the amount of adenine (A) present was always similar to the amount of thymine (T) present and amount of guanine (G) similar to the amount of cytosine (C). This became known as Chargaff’s rule. This, together with the pioneering work of Rosalind Franklin and Maurice Wilkins on the X-ray crystallographic analysis of DNA, led Watson and Crick to believe the pairing of “bases” A+T and G+C was central to the structure of DNA. They used cardboard cutout models of the bases to determine the pattern. At first, nothing seemed to fit, but after Jerry Donohue suggested a more likely tautomeric form of the bases, the jigsaw fell into place. Watson and Crick proposed the now famous double-stranded helix held together

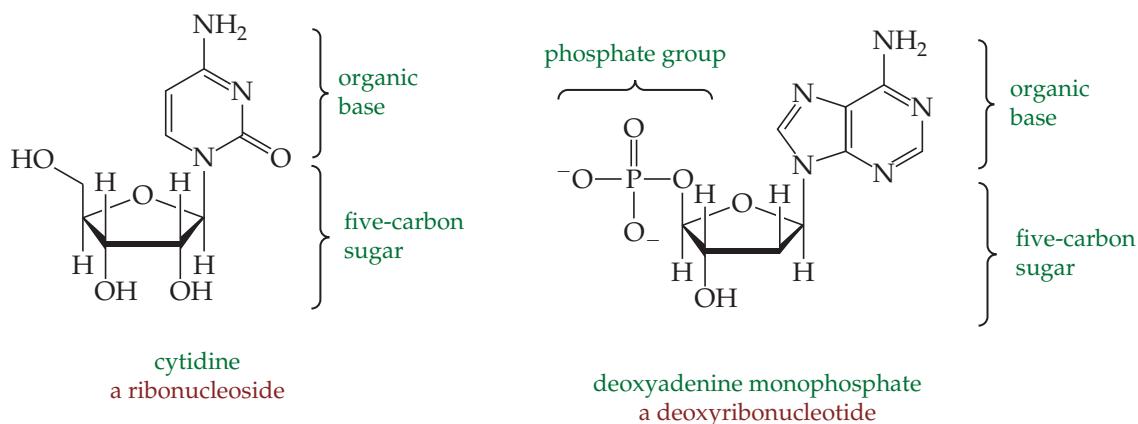
by hydrogen bonding between pairs of bases. They celebrated their discovery with a beer in “The Eagle”, Cambridge, where a plaque commemorates the place where they first announced their discovery to the public.

In this section, we introduce one of the most remarkable known biopolymers. By the end of section, you should:

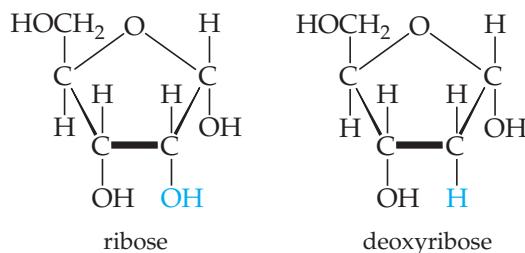
- Understand the components and functioning of DNA.

Nucleic acids make up a class of biopolymers that are the chemical carriers of an organism's genetic information. **Deoxyribonucleic acids (DNA)** are huge molecules whose molecular weights may range from 6 million to 16 million u. **Ribonucleic acids (RNA)** are smaller molecules, with molecular weights in the range of 20000 to 40000 u. DNA is found primarily in the nucleus of the cell, whereas RNA is found mostly outside the nucleus in the *cytoplasm*, the non-nuclear soup enclosed within the cell membrane. DNA stores the genetic information of the cell and specifies which proteins the cell can synthesize. RNA carries the information stored by DNA out of the nucleus of the cell into the cytoplasm as part of the *transcription* process. Once in the cytoplasm, the information, transcribed into messenger RNA (mRNA), is used in protein synthesis. The process of protein synthesis from genetic material, known as *translation*, also involves two other forms of RNA, called ribosomal RNA (rRNA) and transfer RNA (tRNA), which bring together the programmed amino acids in the correct sequence for polypeptide synthesis.

The monomers of nucleic acids can come in two forms. **Nucleosides** are compounds containing a five-carbon sugar and a nitrogen-containing organic base. **Nucleotides** are nucleosides containing an additional phosphate group.



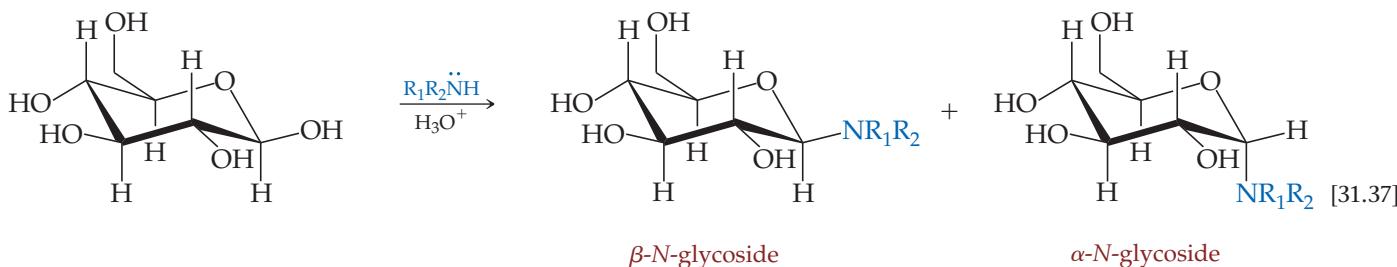
The sugar component of RNA is *ribose*, whereas that in DNA is *deoxyribose*. Deoxyribose differs from ribose only in that it lacks the OH group at carbon 2:



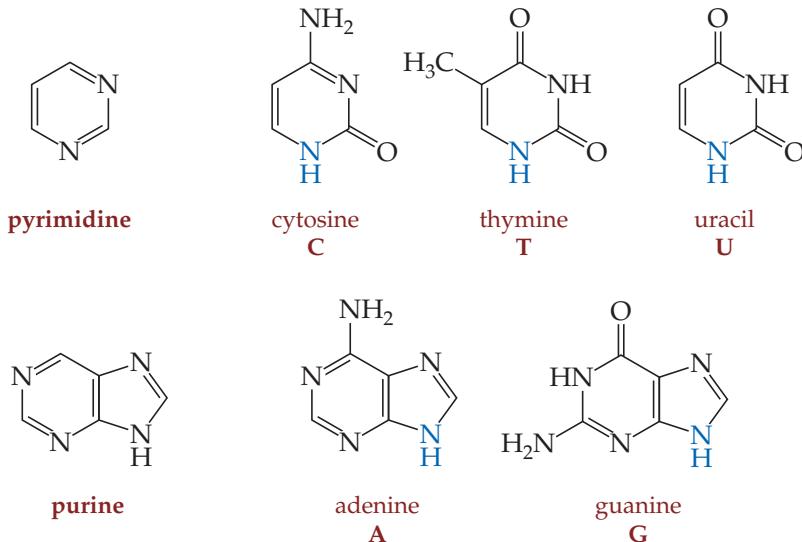
Hence, nucleosides and nucleotides can be either ribose or deoxyribose based. For example, a nucleoside derived from ribose is called a *ribonucleoside* and is found in RNA.

A nucleotide derived from deoxyribose is called a *deoxyribonucleotide* and is found in DNA.

Nucleosides are also known generally as ***N-glycosides***. An *N*-glycoside is formed when a monosaccharide reacts with an amine in the presence of acid. A general reaction for the formation of an *N*-glycoside is



In the case of DNA and RNA, the amines used are derived from the aromatic heterocycles **pyrimidine** and **purine**. The heterocycles adenine (A), thymine (T), cytosine (C) and guanine (G) form the four bases of DNA used to store the genetic code.



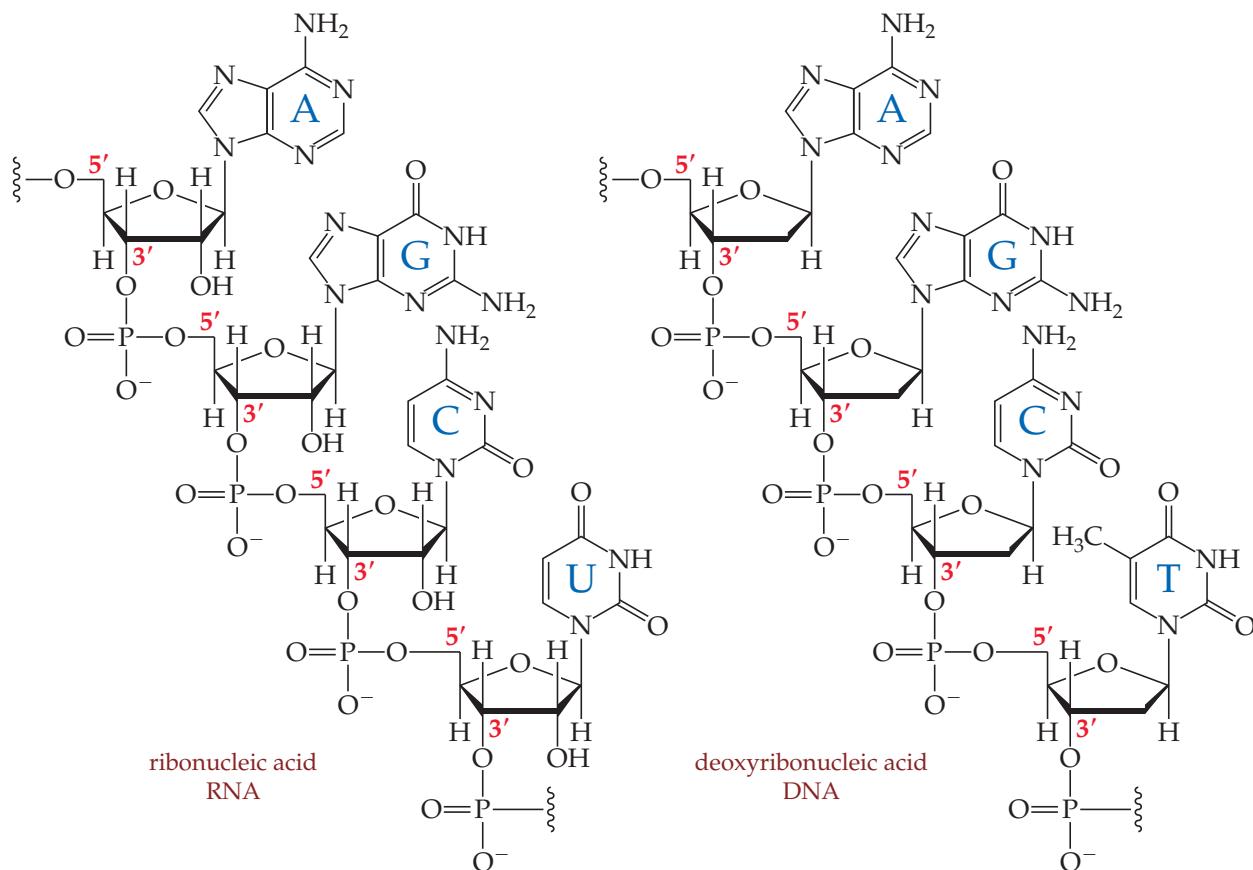
RNA differs from DNA not only in the monosaccharide used, ribose or deoxyribose, but also in the replacement of thymine (T) by uracil (U) within the base code.

Nucleic acids are polynucleotides formed by condensation reactions between an OH group of the phosphoric acid unit at the 5' position on one nucleotide with the OH group at the 3' position of another nucleotide. **Figure 31.28** shows a portion of the polymeric chain of DNA and RNA, illustrating the 3'-5' linkage and the different bases employed by each nucleic acid.

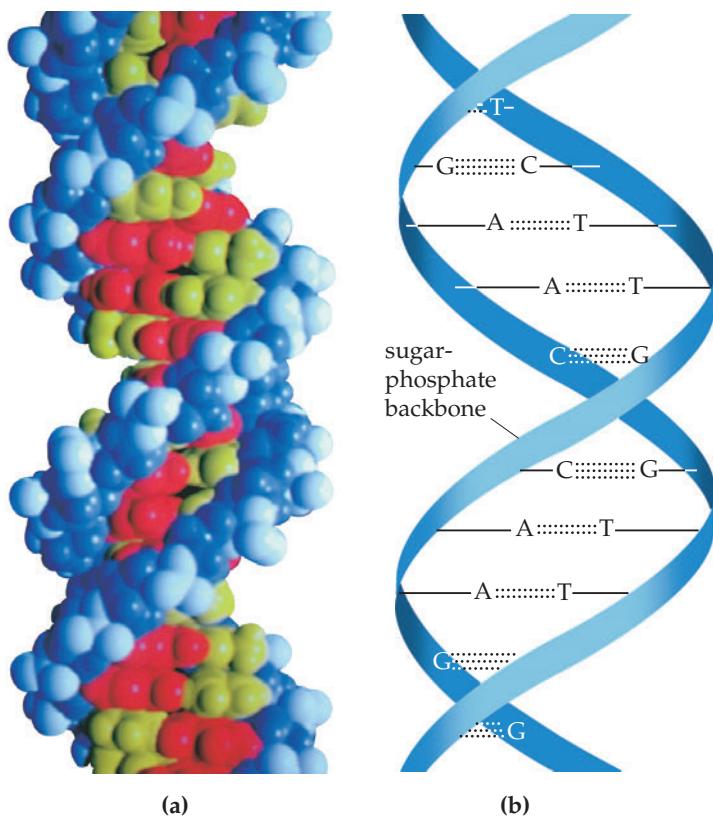
DNA molecules consist of two deoxyribonucleic acid chains or strands that are wound together in the form of a **double helix**, as shown in **Figure 31.29**. The drawing in Figure 31.31(b) has been simplified to show the essential features of the structure. The sugar and phosphate groups form the backbone of each strand. The bases (represented by the letters A, T, C, and G) are attached to the sugars. The two strands are held together by attractions between the bases in one strand and those in the other strand. These attractions involve both London dispersion interactions and hydrogen bonds.

Go Figure

Is DNA positively charged, negatively charged, or neutral in aqueous solution at pH 7?



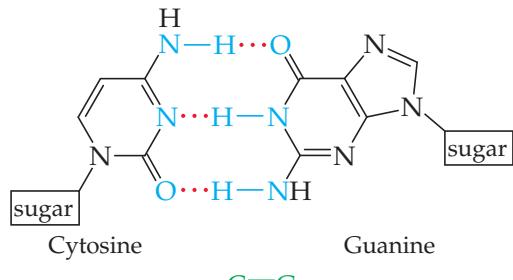
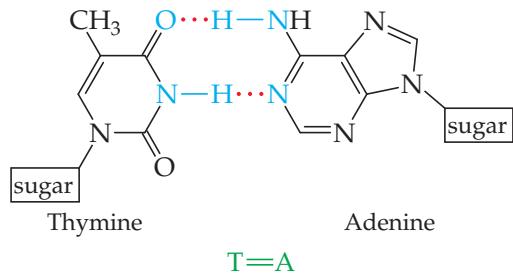
▲ **Figure 31.28** Structure of a polynucleotide. The general structure of DNA and RNA is shown, demonstrating the differences between the two types of nucleic acids.



◀ **Figure 31.29** Two views of DNA. (a) A computer-generated model of a DNA double helix. The dark blue and light blue atoms represent the sugar-phosphate chains that wrap around the outside. Inside the chains are the bases, shown in red and yellow-green. (b) A schematic illustration of the double helix showing the hydrogen bond interactions between complementary base pairs.

 Go Figure

Which pair of complementary bases, A—T or G—C, would you expect to bind more strongly?



▲ **Figure 31.30** Hydrogen bonding between complementary base pairs. The hydrogen bonds are responsible for the formation and stability of the double-stranded helical structure of DNA, shown in Figure 31.29(h).



▲ Figure 31.31 James Watson (1928–) and Francis Crick (1916–2004). Nobel laureates 1962, along with Maurice Wilkins (1916–2004), for their role in the determination of the structure of nucleic acids.

As shown in [Figure 31.30](#), the structures of thymine (T) and adenine (A) make them perfect partners for hydrogen bonding. Likewise, cytosine (C) and guanine (G) form ideal hydrogen bonding partners. In the double-helix structure, therefore, each thymine on one strand is opposite an adenine on the other strand. Likewise, each cytosine is opposite a guanine. This *base-pairing* of C—G and A—T is called **Watson-Crick pairing**, after two of the scientists who were instrumental in solving the structure of DNA ([Figure 31.31](#)). The double-helix structure with complementary bases on the two strands is the key to understanding how DNA functions.

The two strands of DNA unwind during cell division, and new complementary strands are constructed on the unravelling strands (**Figure 31.32**). This enzyme-catalyzed process results in two identical double-helix DNA structures, each containing one strand from the original structure and one newly synthesized strand. This replication process allows genetic information to be transmitted when cells divide.

The structure of RNA and its relationship to DNA is also the key to understanding protein synthesis, the means by which viruses infect cells, and many other problems of central importance to modern biology and medicine. The basis of taking the information stored in DNA and processing it to form proteins is shown in [Figure 31.33](#). The mechanism for protein biosynthesis takes place within a *ribosome*, a complex piece of biological machinery composed of protein and rRNA, which reads the code contained within the mRNA to be translated. This code determines the order of the amino acid sequence within the polypeptide being synthesized.

So how is it possible for just four bases to code for all 20 naturally occurring amino acids? The answer is by using the ribonucleotides, A, C, G and U, in a triplet coding called a **codon**. Each codon is specific for a given amino acid ([Table 31.4](#)). For example, the codon CAU on mRNA codes for histidine to be incorporated into the growing peptide chain, whereas the codon GCA on mRNA codes for alanine to be incorporated into the growing peptide chain.

The reason why translation uses a triplet code is very easy to understand. Using a triplet code based on the four ribonucleotides means that there are $4^3 = 64$ possible triplet

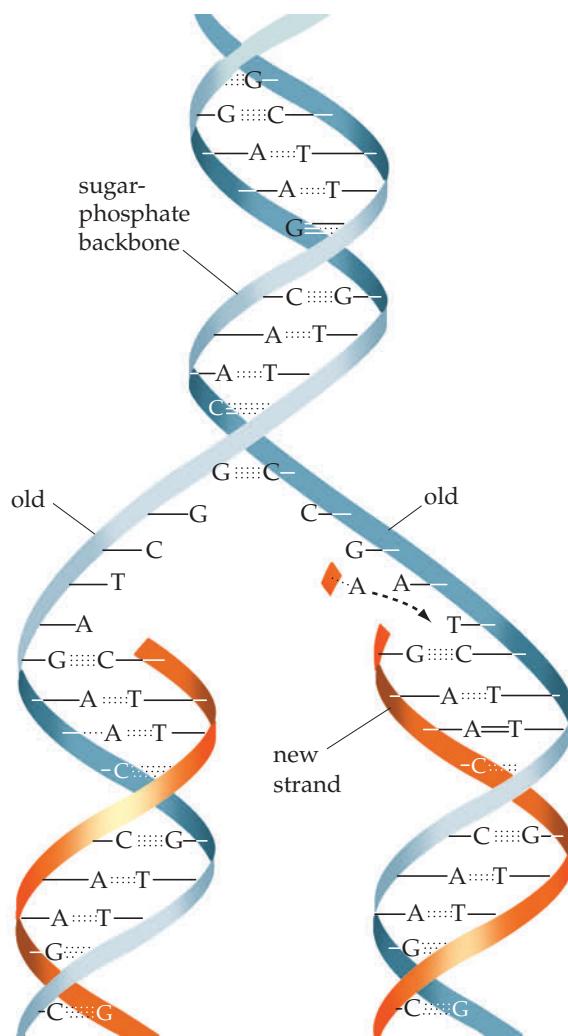


Figure 31.32 DNA replication. The original DNA double helix partially unwinds, and new nucleotides line up on each strand in complementary fashion. Hydrogen bonds help align the new nucleotides with the original DNA chain. When the new nucleotides are joined by condensation reactions, two identical double-helix DNA molecules result.



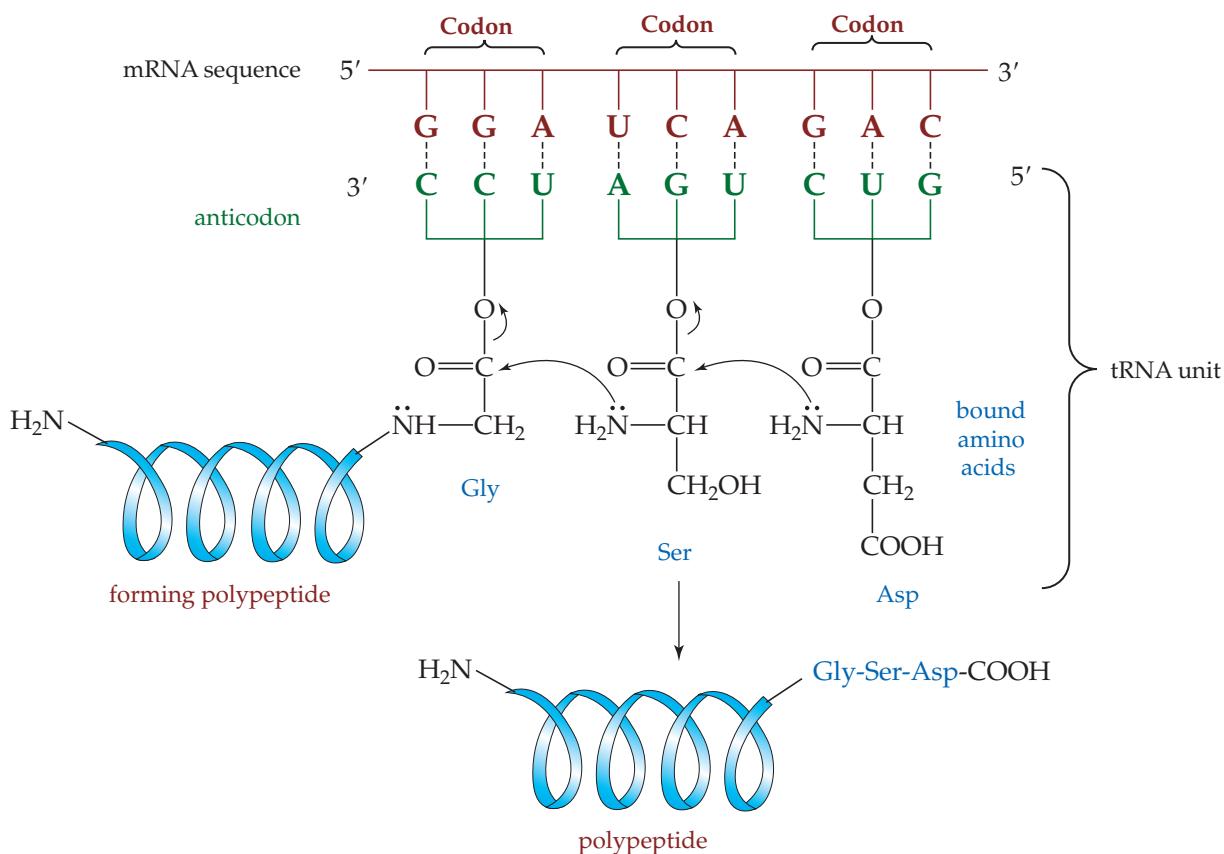
Figure 31.33 From DNA to protein.

codes, which is the smallest number of permutations possible greater than 20, which counts for the 20 naturally occurring amino acids. Of the possible 64 permutations for the codons, 61 are used for specific amino acids. Most amino acids are specified by more than one codon, with the exception of methionine (Met) and tryptophan (Trp). The amino acids serine, leucine and arginine are all specified by any one of six codons. The three remaining codons are used to terminate the synthesis of the polypeptide and release it from the ribosome. These codons are called *stop codons*.

The construction of the polypeptide using the 61 codons sequenced on mRNA requires complementary sequence codons, called *anticodons*, on tRNA. For example, the codon series, 5'-UCA GAC, sequenced on mRNA requires a serine-bound tRNA with the complementary sequence, 3'-AGU-5', as well as an aspartic acid-bound tRNA with the complementary sequence, 3'-CUG-5'. Once bound in the correct orientation, the two units are added to the growing peptide chain by an enzymatic process that cleaves the amino acid from the tRNA anticodon. **Figure 31.34** illustrates the process schematically.

TABLE 31.4 Codon assignments

First base 5' end	Second base	U	C	A	G
Third base 3' end					
U	U	UUU = Phe	UUC = Phe	UUA = Leu	UUG = Leu
	C	UCU = Ser	UCC = Ser	UCA = Ser	UCG = Ser
	A	UAU = Tyr	UAC = Tyr	UAA = Stop	UAG = Stop
	G	UGU = Cys	UGC = Cys	UGA = Stop	UGG = Trp
C	U	CUU = Leu	CUC = Leu	CUA = Leu	CUG = Leu
	C	CCU = Pro	CCC = Pro	CCA = Pro	CCG = Pro
	A	CAU = His	CAC = His	CAA = Gln	CAG = Gln
	G	CGU = Arg	CGC = Arg	CGA = Arg	CGG = Arg
A	U	AUU = Ile	AUC = Ile	AUA = Ile	AUG = Met
	C	ACU = Thr	ACC = Thr	ACA = Thr	ACG = Thr
	A	AAU = Asn	AAC = Asn	AAA = Lys	AAG = Lys
	G	AGU = Ser	AGC = Ser	AGA = Arg	AGG = Arg
G	U	GUU = Val	GUC = Val	GUA = Val	GUG = Val
	C	GCU = Ala	GCC = Ala	GCA = Ala	GCG = Ala
	A	GAU = Asp	GAC = Asp	GAA = Glu	GAG = Glu
	G	GGU = Gly	GGC = Gly	GGA = Gly	GGG = Gly



▲ **Figure 31.34 Translation.** The codons on mRNA orientate the tRNA anticodons bearing amino acid residues. Once bound, enzyme-mediated processes cause the amino acid to be cleaved from the anticodon and covalently bound to the growing peptide chain.



Sample Integrative Problem

Putting Concepts Together

Glycine, $\text{NH}_2\text{CH}_2\text{COOH}$, the simplest of all naturally occurring amino acids, has a melting point of 292°C . The $\text{p}K_a$ of the acid group is 2.35 and the $\text{p}K_a$ associated with the amino group is 9.78. (a) Draw a Lewis structure that indicates the charges on the molecule at the physiological pH of 7.4. (b) Use your structure to illustrate the concept of resonance. (c) Describe the hybridization of the two carbon atoms and the nitrogen atom in glycine and the geometry of the atoms surrounding these three atoms. (d) Glycine has an unusually high melting point for a small molecule. Suggest a reason for this.

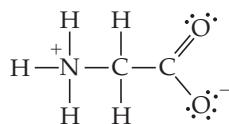
SOLUTION

Analyze We are asked about the structure of a simple amino acid at a particular pH. This includes considering the hybridization and molecular geometry of some of the atoms and commenting on a physical property.

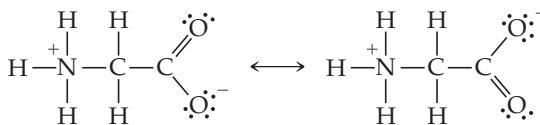
Plan First draw the Lewis structure of glycine and determine the ionization state of the amine and the carboxylic acid groups.

Solve

(a) A pH of 7.4 is several units on the base side of 2.35, consequently the carboxylic acid ($\text{p}K_a = 2.35$) will exist in the conjugate base form. Conversely physiological pH is more than two units on the acid side of 9.78 and so the amine will exist in the conjugate acid form ($\text{p}K_a = 9.78$). The Lewis structure will be:



(b)



(c)

N: sp^3 hybridized, tetrahedral geometry, CH_2 ; sp^3 hybridized, tetrahedral geometry, COO^- ; sp^2 hybridized, trigonal-planar geometry

(d)

Glycine exists as a zwitterion and strong electrostatic attraction between molecules results in some ionic-like characteristics such as a relatively high melting point.

Self-Assessment Exercise

31.30 What codon sequence on mRNA would give the tripeptide: Val-Gly-Gly?

- (a) GUU GGC GGA
(b) UUG GGG GGC

(c) GUC GAA GAG

(d) UGC AGU CAG

Exercises

31.31 Describe a nucleotide. Draw the structural formula for the nucleotide deoxycytidine monophosphate, which contains cytosine in the organic base.

31.32 Imagine a single DNA strand containing a section with the following base sequence: ACTCGA. What is the base sequence of the complementary strand?

31.33 An mRNA strand has the following sequence: GUCAG-GAAUCUU. What would be the sequence of the parent DNA strand that led to this transcription?

31.34 Chargaff's rule states that DNA contains equimolar amounts of guanine and cytosine and also equal amounts of adenine and thymine. (a) Does this imply that there are equimolar amounts of purine and pyrimidine bases in DNA? (b) Does Chargaff's rule apply only to double-stranded DNA or is it also applicable to single-stranded DNA? Explain.

31.35 Poly U added to a cell-free system containing all the necessary materials for biosynthesis yields a polypeptide comprising a single amino acid. (a) What is that amino acid? (b) What synthetic addition polymer might have similar properties to the one formed by translating poly U?

31.36 The motion picture *GATTACA* (1997) is set around a society that strives for human perfection through genetic selection. Its name is a play on the genetic code, comprising the bases A, T, C and G. (a) What is the complementary DNA sequence for GATTACA? (b) Transcribe this sequence into mRNA.

31.37 What amino acid sequence is encoded by the following mRNA sequence?

UUU CCA GUG GAU CCG AUC UAA

31.30 (a)

Answers to Self-Assessment Exercise



Chapter Summary and Key Terms

SECTION 31.1 Amines are hydrocarbon derivatives containing one or more basic nitrogen atoms. **Aliphatic amines** contain alkyl chains. **Heterocycles** containing a basic nitrogen atom are classed as either **cyclic** or **aromatic amines**. **Alkaloids** are amines found in the bark, leaves and roots of plants, for example cocaine, nicotine and morphine. The **amino group** of a primary (RNH_2), secondary ($\text{RR}'\text{NH}$) or tertiary ($\text{RR}'\text{R}''\text{N}$) amine can be converted to an **alkyl ammonium salt** by the addition of acid or alkyl halide. Amines can be synthesized by the **reductive amination** of aldehydes and ketones. The condensation of an amine with a carboxylic acid yields an amide. Amides can be **primary**, **secondary** or **tertiary** based on the number of hydrogen atoms bonded to the amide nitrogen. The amide bond is highly planar due to resonance. **Lactams**, such as those found in the **penicillins**, are cyclic amides.

SECTION 31.2 Compounds containing both carboxylic acid and amine functional groups are called **amino acids**. **α -Amino acids** are chiral substances, with only the L enantiomers found naturally. Since both acidic and basic groups are found within the same molecule, amino acids are said to be **amphiprotic**. At neutral pH, amino acids exist as **zwitterions**. The **isoelectric point** of an amino acid is the pH where the neutral zwitterion predominates in solution. The 20 common naturally occurring amino acids are classed as α -amino acids. They differ only in the identity of their **side chains**. These side chains give the α -amino acids their different physical properties and allow their separation from each other by techniques such as **electrophoresis**. Amino acids can be formed by the **Strecker synthesis**, which converts aldehydes to α -amino acids in the presence of cyanide ion.

SECTION 31.3 Proteins are functional polymers of amino acids. The amino acids within proteins are linked by **peptide (amide) bonds**. A **poly-peptide** is a polymer formed by linking many amino acids by peptide

bonds. A **peptide** is the name given to a short sequence of amino acids. Proteins are classed as either **simple** or **conjugated** based on their constituents. They can be classed also as **fibrous** or **globular** based on their macromolecular shape. Protein structure is determined by the sequence of amino acids in the chain (its **primary structure**), the coiling or stretching of the chain (its **secondary structure**), the overall shape of the complete molecule (its **tertiary structure**) or its interaction with other proteins (its **quaternary structure**). The most important secondary structure arrangements are the **α -helix** and **β -pleated sheet**. The process by which a protein assumes its biologically active tertiary structure is called **folding**. The primary structure of a peptide or protein can be elucidated using the **Edman degradation**. Some proteins, called **enzymes**, function as catalysts. Enzymes contain an **active site**, which is the area of reactivity for a **substrate**. The active site and substrate form a very specific and functioning complex, much like a **lock and key**.

SECTION 31.4 Nucleic acids are biopolymers that carry the genetic information necessary for cell reproduction and development through control of protein synthesis. The building blocks of these biopolymers are **nucleosides** and phosphoric acid, which react to form **nucleotides**. Nucleosides are also generally known as **N-glycosides**. There are two types of nucleic acids: the **ribonucleic acids (RNA)** and the **deoxyribonucleic acids (DNA)**. These substances consist of a polymeric backbone of alternating phosphate and sugar groups, with **pyrimidine**- or **purine**-derived organic bases attached to the sugars. The DNA forms a **double helix** held together by hydrogen bonding between matching organic bases on the two strands. The specific nature of **Watson-Crick pairing** is the key to gene replication and protein synthesis. Each set of three nucleotides, or **codon**, within messenger RNA (mRNA) codes for a particular amino acid. This process, called **translation**, converts the genetic code into protein.

Key Skills

- Be able to name and classify amines and amides. (Section 31.1)
- Recall that amines are both bases and nucleophiles and appreciate relative pK_a values. (Section 31.1)
- Prepare amines and amides from other functional groups. (Section 31.1)
- Know the 20 most common amino acids and their three-letter codes. Expand this to describe small peptides. (Section 31.2)
- Be able to use the Henderson-Hasselbalch equation. (Section 31.2)
- Have an understanding of the importance to biological processes of nitrogen-containing compounds. (Sections 31.3 and 31.4)

- Know the general classes of proteins and describe what is meant by primary, secondary, tertiary and quaternary structure of a protein. (Section 31.3)
- Be able to sequence a peptide through the fragment approach. (Section 31.3)
- Identify the purine and pyrimidine bases used in nucleic acids. (Section 31.4)
- Be able to show A/T and C/G complementarity in a DNA/DNA or RNA/DNA sequence. (Section 31.4)
- Have an understanding of the process of translation and transcription. (Section 31.4)

Key Equations

- Isoelectric point

$$\text{pI} = \frac{1}{2} (\text{p}K_{\text{a}1} + \text{p}K_{\text{a}2}) \quad [31.23]$$

- Henderson-Hasselbalch equation

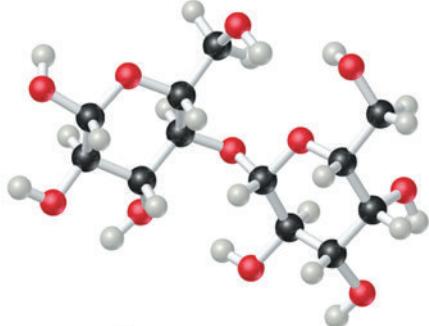
$$\text{pH} = \text{p}K_{\text{a}} + \log \frac{[\text{conjugate base}]}{[\text{weak acid}]} \quad [31.28]$$

Exercises

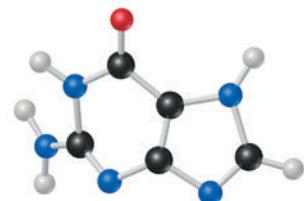
Visualizing Concepts

- 31.38** From examination of the following ball-and-stick molecular models, choose the substance that **(a)** can be hydrolyzed to form a solution containing glucose, **(b)** is capable of forming a zwitterion, **(c)** is one of the four bases present in DNA. [Sections 31.2 and 31.4]

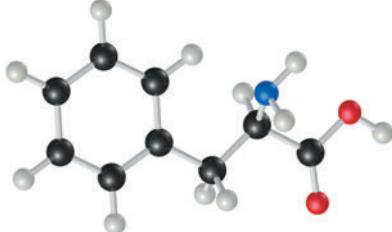
(i)



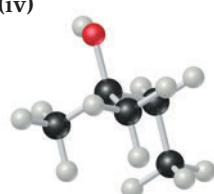
(ii)



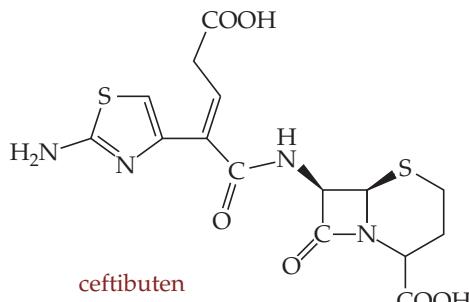
(iii)



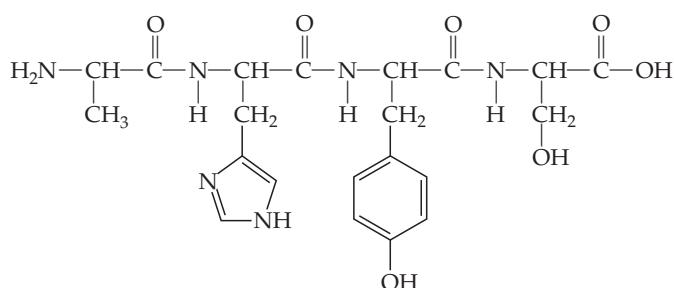
(iv)



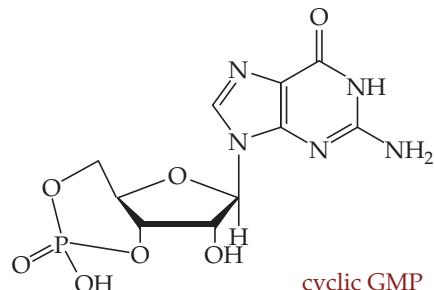
- 31.39** Ceftibuten is an orally active cephalosporin antibiotic. Identify the β -lactam, secondary amide, primary amine and heterocyclic amine in this structure. [Section 31.1]



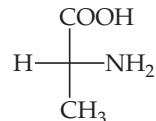
- 31.40** What are the three-letter and one-letter codes for the following tetrapeptide? [Section 31.3]



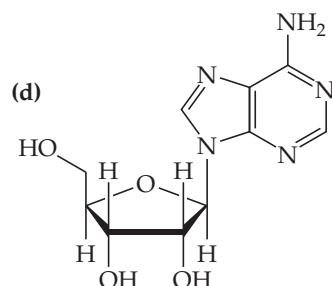
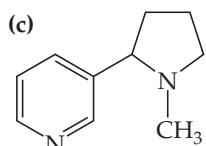
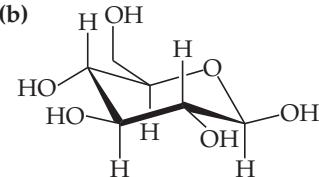
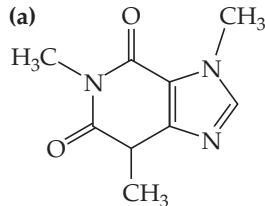
- 31.41** Cyclic GMP is a cellular regulatory agent found in animal and bacteria cells in concentrations of 10^{-6} moles kg^{-1} . Cyclic GMP is also implicated in the action and potency of Viagra. **(a)** Can you suggest what GMP might stand for? **(b)** What is the name given to the monosaccharide central to the structure of cyclic GMP? **(c)** Can cyclic GMP be classified as a glycoside? If so, what type of glycoside? [Section 31.4]

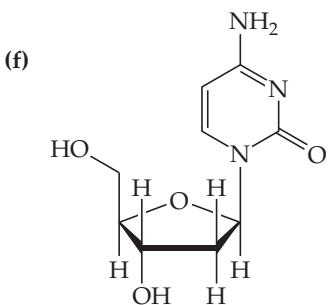
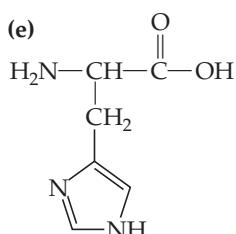


- 31.42** The configuration of an amino acid is drawn here as a Fischer projection. **(a)** Name the amino acid. **(b)** Assign its absolute stereochemistry as *R* or *S*. [Section 31.2]

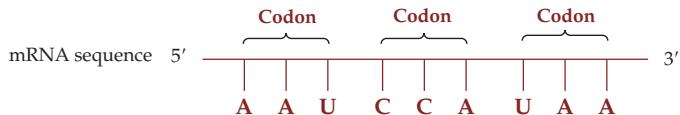


- 31.43** Classify the following compounds as an alkaloid, amino acid, nucleoside, or carbohydrate. [Sections 31.1, 31.2 and 31.4]





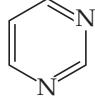
- 31.44** Identify the peptide formed by the mRNA sequence following translation. [Section 31.4]



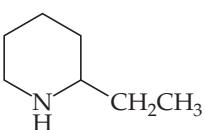
Amines and the Amide bond (Section 31.1)

- 31.45** Give names for the following amines:

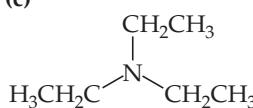
(a)



(b)



(c)



(d)



- 31.46** Provide a structural formula for each of the following compounds:

(a) 2-propanamine

(b) 2-aminobutanoic acid

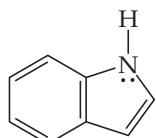
(c) *N,N*-dimethylcyclopentanamine

(d) 2-methylcyclohexylamine

(e) 2-methyl-3-ethylpyrrole

- 31.47** (a) Draw a primary, secondary, and tertiary amine with the molecular formula $\text{C}_6\text{H}_{15}\text{N}$. What relationship exists between the three compounds you have drawn? (b) How would you expect the boiling points of these three amines to vary, and why? (c) Propylamine is entirely miscible with water; trimethylamine has reasonably high solubility in water. What accounts for these data, considering that isobutane ($(\text{CH}_3)_3\text{CH}$) is considerably less soluble than trimethylamine?

- 31.48** Indole smells rather terrible in high concentrations but has a pleasant floral odor when highly diluted. It has the following structure:



Indole is a planar molecule. The nitrogen is a very weak base, with a K_b of 2×10^{-12} . Explain how the K_b indicates that the indole molecule is aromatic in character.

- 31.49** Aniline is prepared by catalytic hydrogenation of nitrobenzene. (a) Write an equation for this reaction including reagents and catalyst. (b) Devise a chemical procedure based on the basicity of aniline to separate it from any unreacted nitrobenzene. [Hint: Both compounds are soluble in diethyl ether and insoluble in water.]

- 31.50** Provide a structural formula for each of the following compounds:

(a) 2-methylpropanamide

(b) acetamide

(c) *N*-phenylacetamide

(d) *N,N*-dimethyl-1-pentanamide

(e) *N*-ethylbenzamide

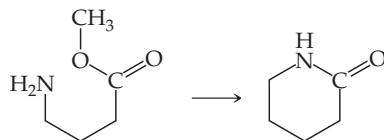
- 31.51** Draw a primary, secondary and tertiary amide with the molecular formula $\text{C}_5\text{H}_{11}\text{NO}$. What isomeric relationship exists between the three compounds you have drawn?

- 31.52** *Capsaicin*, shown in the accompanying figure, is the compound responsible for the sensation felt when chillis are eaten. It is often used as a tool in neurobiological research and has application as a topical analgesic. (a) Draw the structure of the carboxylic acid and amine that lead to capsaicin. (b) What is the general name given to this type of reaction?

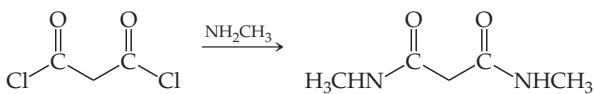


- 31.53** Propose a mechanism for the following reactions:

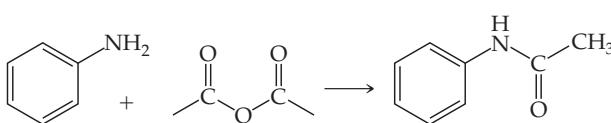
(a)



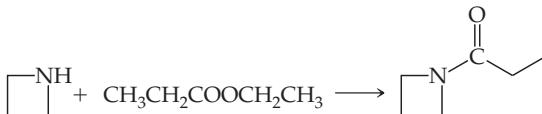
(b)



(c)

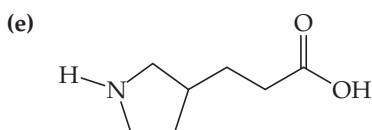
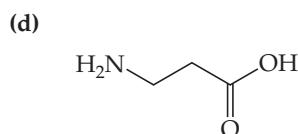
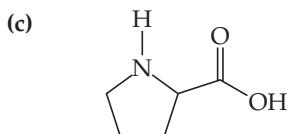
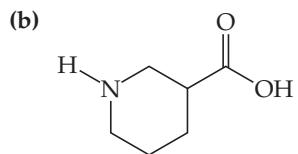
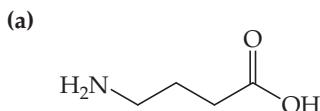


(d)

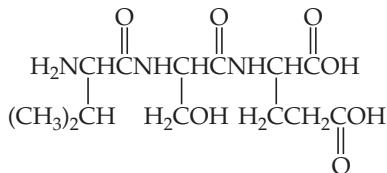


Amino Acids, Proteins, and Peptides (Sections 31.2 and 31.3)

- 31.54** What properties of the side chains (R groups) of amino acids are important in affecting the amino acid's overall biochemical behaviour? Give examples to illustrate your answer.
- 31.55** Classify the following five molecules as α -, β -, γ - or δ -amino acids:



- 31.56** Aspartic acid is reported to have three pK_a values, one in the range of 2 to 3, another in the range 3 to 4, and the last in the range of 9 to 10. Suggest the origin of the three pK_a values.
- 31.57** (a) What is meant by the term *isoelectric point*? (b) Valine has pK_a values of 2.32 and 9.69. Determine the isoelectric point, pI , for valine.
- 31.58** Using Equation 31.28, estimate the average charge on phenylalanine (a) at pH 1, (b) at pH 5.5, (c) at pH 8. Phenylalanine has pK_a values of 2.58 and 9.24.
- 31.59** (a) Draw leucine and lysine in their zwitterionic forms. (b) The pI for leucine and lysine are 5.98 and 9.74, respectively. How could you use this information to separate a mixture of the two amino acids by electrophoresis?
- 31.60** Outline a synthesis of isoleucine (a) from 2-bromo-3-methylpentanoic acid and (b) by the Strecker synthesis.
- 31.61** (a) Write a chemical equation for the formation of methionyl glycine from its constituent amino acids. (b) If a racemic mix of both amino acids is used, how many stereoisomers will result?
- 31.62** (a) What amino acids would be obtained by hydrolysis of the following tripeptide?



- (b) How many different tripeptides can be made from the amino acids glycine, serine and glutamic acid? Give the abbreviation for each of these tripeptides, using the three-letter codes for amino acids.

- 31.63** (a) Describe the role of hydrogen bonding in determining the α -helical structure of a protein. (b) Suggest a reason why proline is never found within a protein α -helix.

- 31.64** (a) Draw the condensed structural formula of each of these tripeptides: Val-Gly-Asp, Phe-Ser-Ala. (b) What general properties would you expect each tripeptide to have based on their side chains—for example, are they hydrophobic or hydrophilic, acidic or basic?

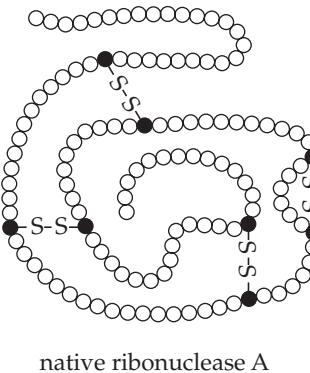
- 31.65** Draw the full structural formula for alanylglycylserine.

- 31.66** A new cyclic pentapeptide has been isolated from a marine sponge. Total hydrolysis of the cyclic peptide yielded the following amino acid composition:

Lys	Gly	Phe	Cys
-----	-----	-----	-----

- (a) What experimental conditions and reagents could be used to perform this hydrolysis? (b) What techniques could be used to cleave the cyclic peptide at specific sites to help identify the structure?

- 31.67** The protein ribonuclease A in its native, or most stable, form is folded into a compact globular shape. (a) Does the native form have a lower or higher free energy than the denatured form, in which the protein is an extended chain? (b) What is the sign of the entropy change in going from the denatured to the folded form? (c) In its folded form, the ribonuclease A has four $-S-S-$ bonds that bridge parts of the chain, as shown in the diagram here. What effect do you predict that these four linkages have on the free energy and entropy of the folded form as compared with a hypothetical folded structure that does not possess the four $-S-S-$ linkages? Explain. (d) A gentle reducing agent converts the four $-S-S-$ linkages to eight $-S-H$ bonds. What effect do you predict this would have on the tertiary structure of the protein?

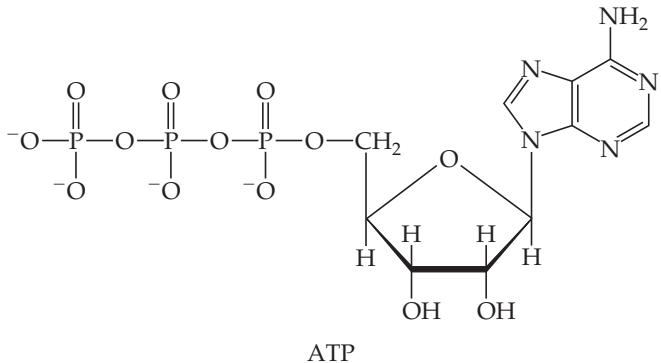


Nucleic Acids and DNA (Section 31.4)

- 31.68** A nucleoside consists of an organic base of the kind shown in Section 31.4, bound to ribose or deoxyribose. Draw the structure for deoxyguanosine, formed from guanine and deoxyribose.

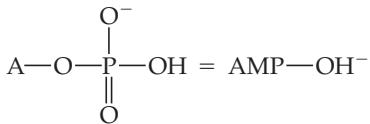
- 31.69** When samples of double-stranded DNA are analysed, the molar quantity of adenine present equals that of thymine. Similarly, the molar quantity of guanine equals that of cytosine. Explain the significance of these observations.

- 31.70** One of the most important molecules in biochemical systems is adenosine triphosphate (ATP), for which the structure is:



ATP is the principal carrier of biochemical energy. It is considered an energy-rich compound because the hydrolysis of ATP to yield adenosine diphosphate (ADP) and inorganic phosphate is spontaneous under aqueous biochemical conditions. **(a)** Write a balanced equation for the reaction of ATP with water to yield ADP and inorganic phosphate ion. **(b)** What would you expect for the sign of the free energy change for this reaction? **(c)** ADP can undergo further hydrolysis. What would you expect for the product of that reaction?

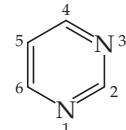
- 31.71** The monoanion of adenosine monophosphate (AMP) is an intermediate in phosphate metabolism:



where A = adenine. If the pK_a for this anion is 7.21, what is the ratio of $[\text{AMP}-\text{OH}^-]$ to $[\text{AMP}-\text{O}^{2-}]$ in blood at pH 7.4?

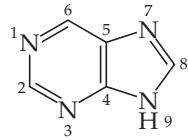
- 31.72** **(a)** Draw the structure of adenine and thymine and indicate how they interact through Watson–Crick base pairing. **(b)** Draw the structure of cytosine and guanine and indicate how they interact through Watson–Crick base pairing. **(c)** Why doesn't guanine base-pair with thymine? Explain.

- 31.73** *Broxuridine* is used to replace thymine during replication to cause radiosensitivity. Its systematic name is 5-bromo-2'-deoxyuridine or 5-bromouracil deoxyriboside. Draw the structural formula for broxuridine.



pyrimidine

- 31.74** *Cladribine* is a substituted purine nucleoside with anti-leukaemic activity. Its systematic name is 2-chloro-2'-deoxyadenosine. Draw the structural formula for cladribine.



purine

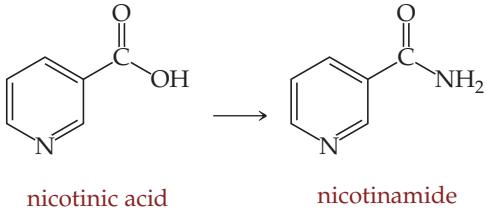
- 31.75** **(a)** What amino acid sequence is encoded by the following mRNA sequence?

AGG ACU GCA UCG CAA

- (b)** What anticodon sequence of tRNAs is needed for the synthesis of this peptide?

Integrative Exercises

- 31.76** Devise a synthesis for nicotinamide starting from nicotinic acid.



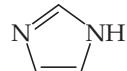
- 31.77** **(a)** How many amide groups are there in vitamin B₁₂? See page 1436. **(b)** What type of glycoside is found in vitamin B₁₂? **(c)** Is the sugar component a furanose or pyranose form?

- 31.78** What general features of a vitamin allow you to predict whether it is a hydrophilic or a lipophilic vitamin?

- 31.79** Quinine is a natural product that has anti-malarial properties. It was originally extracted for therapeutic use from the bark of the cinchona tree, but is now synthesized

by the pharmaceutical industry. Quinine is generally administered as the hydrochloride salt ($\text{C}_{20}\text{H}_{24}\text{N}_2\text{O}_2 \cdot \text{HCl}$). The pK_a of this salt is 4.32. **(a)** What are the advantages of using the hydrochloride salt of quinine in medicine? **(b)** What is the pH of a 0.053 M solution of quinine hydrochloride?

- 31.80** Imidazole is a cyclic amine with the following structure.



- (a)** Draw a tautomer of imidazole. **(b)** Imidazole is a highly polar molecule; draw a resonance structure that demonstrates this polarity. **(c)** Imidazole is aromatic with six π -electrons; identify the π -electrons involved. **(d)** Protonation of imidazole occurs on only one of the nitrogen atoms. Which nitrogen atom is protonated and explain why this is so? Imidazole forms part of the side chain of the amino acid histidine and coordinates to iron in the oxygen transport protein, hemoglobin. **(e)** What part of the imidazole molecule acts as a donor in this coordination?

Design an Experiment

You are tasked with identifying the sequence of a peptide in the most efficient way possible. You conduct a series of experiments and collect the following data:

- Edman degradation on the peptide yields Cys.
- Chymotrypsin hydrolysis gives two fragments; one fragment contains the amino acids Cys and Lys and the other fragment contains Cys, Gly, and Phe.
- Trypsin hydrolysis also gives two fragments; one containing the amino acids Cys, Gly, Lys, and Phe and another composed of just Cys.

What is the sequence of the peptide? If you had repeated the Edman degradation until you had identified each of the amino acids present, how many experiments would this have involved? If you had conducted an acid hydrolysis to release all of the amino acids in a single experiment, can you think of a way to quantify the amino acids present?