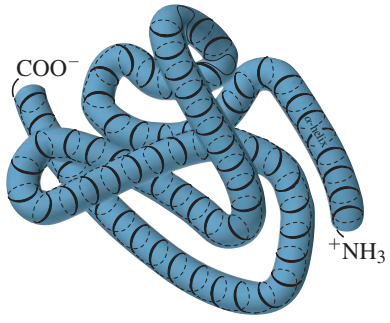
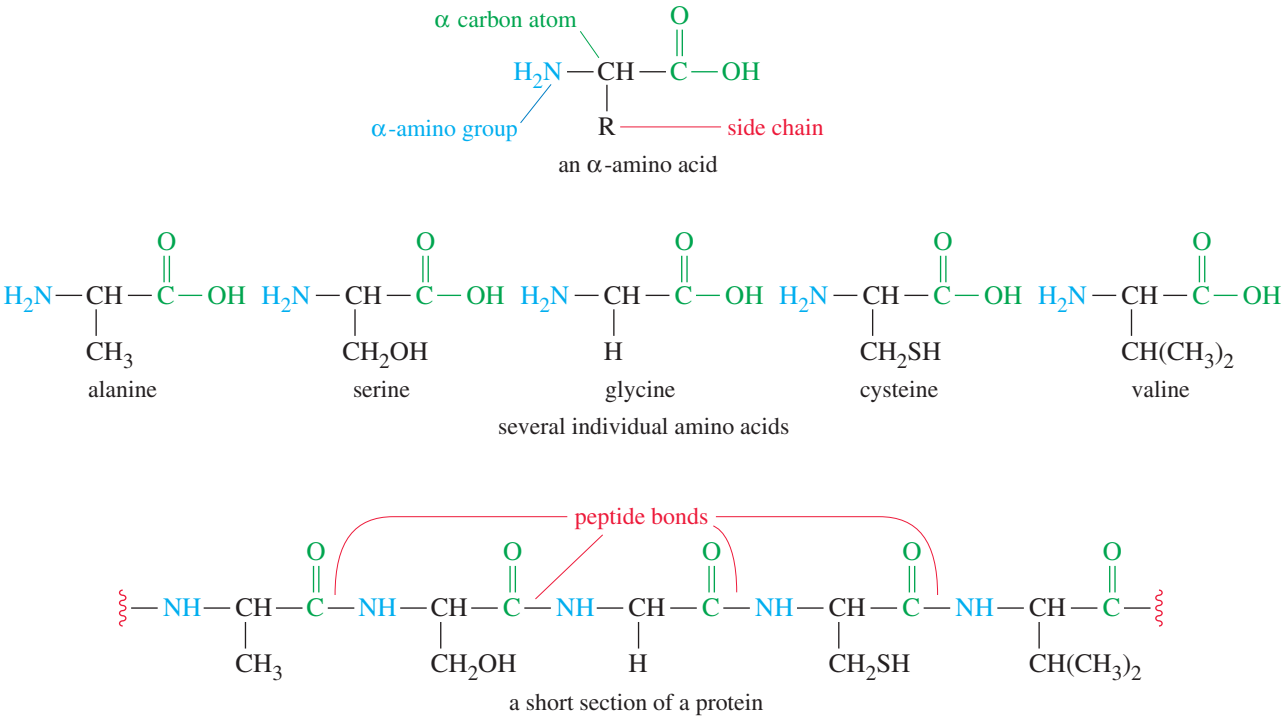
CHAPTER

# 24 AMINO ACIDS, PEPTIDES, AND PROTEINS




## 24-1 Introduction

Proteins are the most abundant organic molecules in animals, playing important roles in all aspects of cell structure and function. Proteins are biopolymers of **α-amino acids**, so named because the amino group is bonded to the α carbon atom, next to the carbonyl group. The physical and chemical properties of a protein are determined by its constituent amino acids. The individual amino acid subunits are joined by amide linkages called **peptide bonds**. Figure 24-1 shows the general structure of an α-amino acid and a protein.




■ **FIGURE 24-1**
Structure of a general protein and its constituent amino acids. The amino acids are joined by amide linkages called peptide bonds.

**TABLE 24-1**

Examples of Protein Functions

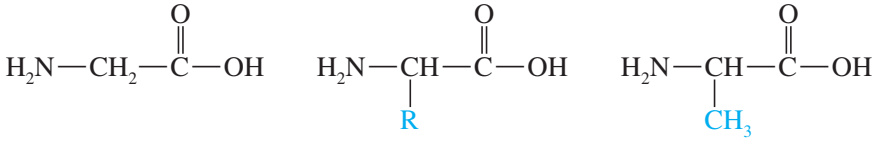
| Class of Protein | Example | Function of Example |
|---|---|---|
| structural proteins | collagen, keratin | strengthen tendons, skin, hair, nails |
| enzymes | DNA polymerase | replicates and repairs DNA |
| transport proteins | hemoglobin | transports $O_2$ to the cells |
| contractile proteins | actin, myosin | cause contraction of muscles |
| protective proteins | antibodies | complex with foreign proteins |
| hormones | insulin | regulates glucose metabolism |
| toxins | snake venoms | incapacitate prey |

Proteins have an amazing range of structural and catalytic properties as a result of their varying amino acid composition. Because of this versatility, proteins serve an astonishing variety of functions in living organisms. Some of the functions of the major classes of proteins are outlined in Table 24-1.

The study of proteins is one of the major branches of biochemistry, and there is no clear division between the organic chemistry of proteins and their biochemistry. In this chapter, we begin the study of proteins by learning about their constituents, the amino acids. We also discuss how amino acid monomers are linked into the protein polymer, and how the properties of a protein depend on those of its constituent amino acids. These concepts are needed for the further study of protein structure and function in a biochemistry course.

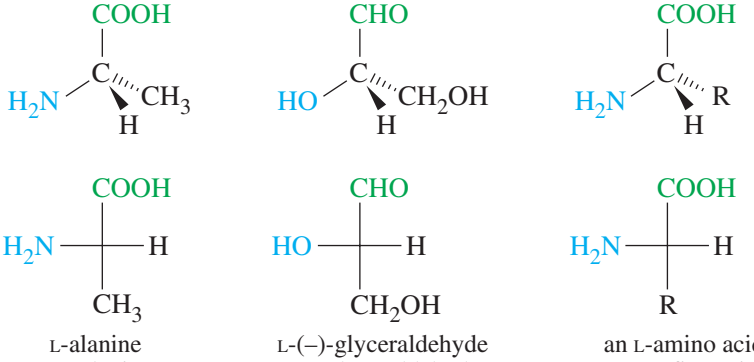
## 24-2 Structure and Stereochemistry of the α-Amino Acids

The term **amino acid** might mean any molecule containing both an amino group and any type of acid group; however, the term is almost always used to refer to an α-amino carboxylic acid. The simplest α-amino acid is aminoacetic acid, called *glycine*. Other common amino acids have side chains (symbolized by R) substituted on the α carbon atom. For example, alanine is the amino acid with a methyl side chain.




glycine | a substituted amino acid | alanine (R = $CH_3$)

Except for glycine, the α-amino acids are all chiral. In all of the chiral amino acids, the chirality center is the asymmetric α carbon atom. Nearly all the naturally occurring amino acids are found to have the *(S)* configuration at the α carbon atom. Figure 24-2 shows a Fischer projection of the *(S)* enantiomer of alanine, with the carbon chain along the vertical and the carbonyl carbon at the top. Notice that the configuration of *(S)*-alanine is similar to that of L-(−)-glyceraldehyde, with the amino group on the left in the Fischer




L-alanine *(S)*-alanine | L-(−)-glyceraldehyde *(S)*-glyceraldehyde | an L-amino acid *(S)* configuration

**FIGURE 24-2**
Almost all the naturally occurring amino acids have the *(S)* configuration. They are called L-amino acids because their stereochemistry resembles that of L-(−)-glyceraldehyde.

projection. Because their stereochemistry is similar to that of L-(−) -glyceraldehyde, the naturally occurring *(S)*-amino acids are classified as **L-amino acids**.

Although D-amino acids are occasionally found in nature, we usually assume the amino acids under discussion are the common L-amino acids. Remember once again that the D and L nomenclature, like the *R* and *S* designation, gives the configuration of the asymmetric carbon atom. It does not imply the sign of the optical rotation, (+) or (−), which must be determined experimentally.

Amino acids combine many of the properties and reactions of both amines and carboxylic acids. The combination of a basic amino group and an acidic carboxyl group in the same molecule also results in some unique properties and reactions. The side chains of some amino acids have additional functional groups that lend interesting properties and undergo reactions of their own.

Bacteria require specific enzymes, called racemases, to interconvert D and L amino acids. Mammals do not use D amino acids, so compounds that block racemases do not affect mammals and show promise as antibiotics.

## 24-2A The Standard Amino Acids of Proteins

The **standard amino acids** are 20 common α -amino acids that are found in nearly all proteins. The standard amino acids differ from each other in the structure of the side chains bonded to their α carbon atoms. All the standard amino acids are L-amino acids. Table 24-2 shows the 20 standard amino acids, grouped according to the

**TABLE 24-2**

The Standard Amino Acids

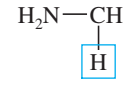
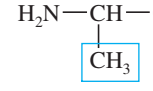
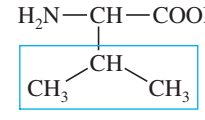
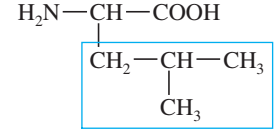
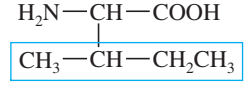
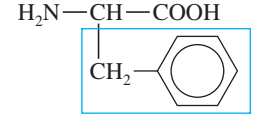
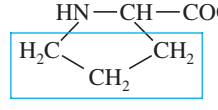
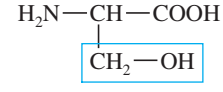
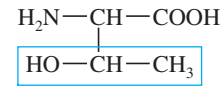
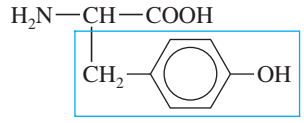
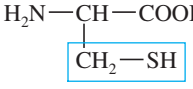
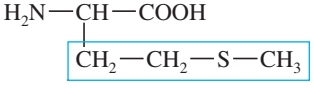
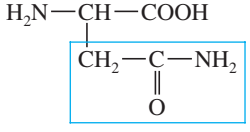
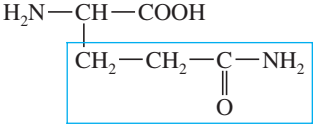
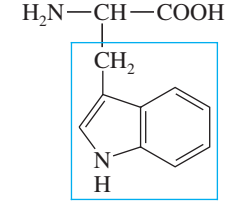
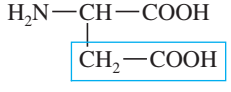
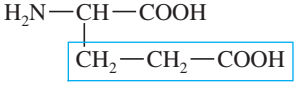
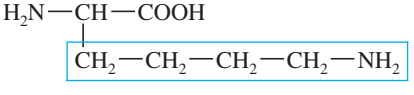
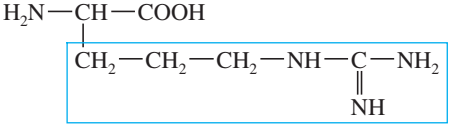
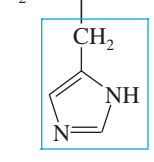
| Name | Symbol | Abbreviation | Structure | Functional Group in Side Chain | Isoelectric Point |
|---|---|---|---|---|---|
| side chain is nonpolar, H or alkyl | | | | | |
| glycine | G | Gly |  | none | 6.0 |
| alanine | A | Ala |  | alkyl group | 6.0 |
| *valine | V | Val |  | alkyl group | 6.0 |
| *leucine | L | Leu |  | alkyl group | 6.0 |
| *isoleucine | I | Ile |  | alkyl group | 6.0 |
| *phenylalanine | F | Phe |  | aromatic group | 5.5 |
| proline | P | Pro |  | rigid cyclic structure | 6.3 |
| side chain contains an —OH | | | | | |
| serine | S | Ser |  | hydroxyl group | 5.7 |
| *threonine | T | Thr |  | hydroxyl group | 5.6 |

**TABLE 24-2**

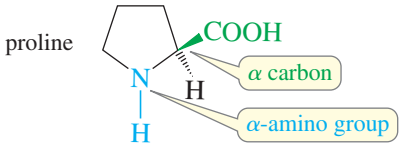
The Standard Amino Acids (*continued*)

| Name | Symbol | Abbreviation | Structure | Functional Group in Side Chain | Isoelectric Point |
|---|---|---|---|---|---|
| tyrosine | Y | Tyr |  | phenolic —OH group | 5.7 |
| **side chain contains sulfur** | | | | | |
| cysteine | C | Cys |  | thiol | 5.0 |
| *methionine | M | Met |  | sulfide | 5.7 |
| **side chain contains nonbasic nitrogen** | | | | | |
| asparagine | N | Asn |  | amide | 5.4 |
| glutamine | Q | Gln |  | amide | 5.7 |
| *tryptophan | W | Trp |  | indole | 5.9 |
| **side chain is acidic** | | | | | |
| aspartic acid | D | Asp |  | carboxylic acid | 2.8 |
| glutamic acid | E | Glu |  | carboxylic acid | 3.2 |
| **side chain is basic** | | | | | |
| *lysine | K | Lys |  | amino group | 9.7 |
| *arginine | R | Arg |  | guanidino group | 10.8 |
| *histidine | H | His |  | imidazole ring | 7.6 |

*essential amino acid

chemical properties of their side chains. Each amino acid is given a three-letter abbreviation and a one-letter symbol (green) for use in writing protein structures.

Notice in Table 24-2 how proline is different from the other standard amino acids. Its amino group is fixed in a ring with its $\alpha$ carbon atom. This cyclic structure lends additional strength and rigidity to proline-containing peptides.




**PROBLEM 24-1**

Draw three-dimensional representations of the following amino acids.
**(a)** L-phenylalanine **(b)** L-histidine **(c)** D-serine **(d)** L-tryptophan

**PROBLEM 24-2**

Most naturally occurring amino acids have chirality centers (the asymmetric $\alpha$ carbon atoms) that are named *(S)* by the Cahn–Ingold–Prelog convention (Section 5-3). The common naturally occurring form of cysteine has a chirality center that is named *(R)*, however.
**(a)** What is the relationship between *(R)*-cysteine and *(S)*-alanine? Do they have the opposite three-dimensional configuration (as the names might suggest) or the same configuration?
**(b)** *(S)*-alanine is an L-amino acid (Figure 24-2). Is *(R)*-cysteine a D-amino acid or an L-amino acid?

## 24-2B Essential Amino Acids

Humans can synthesize about half of the amino acids needed to make proteins. Other amino acids, called the **essential amino acids**, must be provided in the diet. The ten essential amino acids, starred (*) in Table 24-2, are the following:

| | | | |
|---|---|---|---|
| arginine (Arg) | valine (Val) | methionine (Met) | leucine (Leu) |
| threonine (Thr) | phenylalanine (Phe) | histidine (His) | isoleucine (Ile) |
| lysine (Lys) | tryptophan (Trp) | | |

Proteins that provide all the essential amino acids in about the right proportions for human nutrition are called **complete proteins**. Examples of complete proteins are those in meat, fish, milk, and eggs. About 50 g of complete protein per day is adequate for adult humans.

Proteins that are severely deficient in one or more of the essential amino acids are called **incomplete proteins**. If the protein in a person's diet comes mostly from one incomplete source, the amount of human protein that can be synthesized is limited by the amounts of the deficient amino acids. Plant proteins are generally incomplete. Rice, corn, and wheat are all deficient in lysine. Rice also lacks threonine, and corn also lacks tryptophan. Beans, peas, and other legumes have the most complete proteins among the common plants, but they are deficient in methionine.

Vegetarians can achieve an adequate intake of the essential amino acids if they eat many different plant foods. Plant proteins can be chosen to be complementary, with some foods supplying amino acids that others lack. An alternative is to supplement the vegetarian diet with a rich source of complete protein such as milk or eggs.

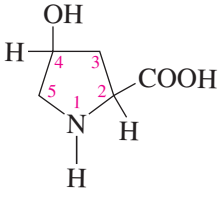
Gelatin is made from collagen, which is a structural protein composed primarily of glycine, proline, and hydroxyproline. As a result, gelatin has low nutritional value because it lacks many of the essential amino acids.

**PROBLEM 24-3**

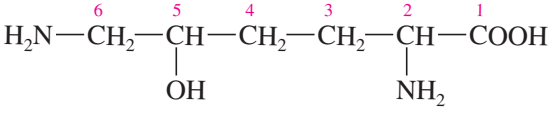
The herbicide *glyphosate* (Roundup®) kills plants by inhibiting an enzyme needed for synthesis of phenylalanine. Deprived of phenylalanine, the plant cannot make the proteins it needs, and it gradually weakens and dies. Although a small amount of glyphosate is deadly to a plant, its human toxicity is quite low. Suggest why this powerful herbicide has little effect on humans.

### 24-2C Rare and Unusual Amino Acids

In addition to the standard amino acids, other amino acids are found in protein in smaller quantities. For example, 4-hydroxyproline and 5-hydroxylysine are hydroxylated versions of standard amino acids. These are called *rare* amino acids, even though they are commonly found in collagen.

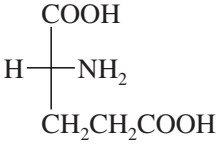


4-hydroxyproline

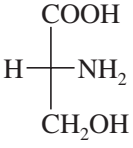


5-hydroxylysine

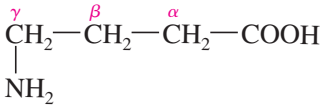
Some of the less common D enantiomers of amino acids are also found in nature. For example, D-glutamic acid is found in the cell walls of many bacteria, and D-serine is found in earthworms. Some naturally occurring amino acids are not $\alpha$-amino acids: $\gamma$-Aminobutyric acid (GABA) is one of the neurotransmitters in the brain, and $\beta$-alanine is a constituent of the vitamin pantothenic acid.



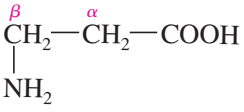
D-glutamic acid



D-serine



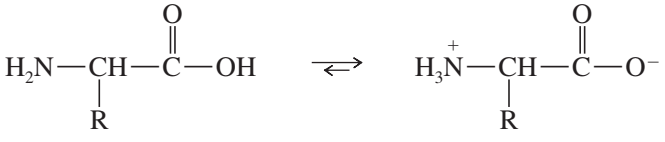
$\gamma$-aminobutyric acid



$\beta$-alanine

## 24-3 Acid–Base Properties of Amino Acids

Although we commonly write amino acids with an intact carboxyl (—COOH) group and amino ($—NH_2$) group, their actual structure is ionic and depends on the pH. The carboxyl group loses a proton, giving a carboxylate ion, and the amino group is protonated to an ammonium ion. This structure is called a **dipolar ion** or a **zwitterion** (German for "dipolar ion").

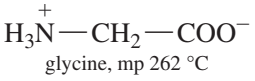


uncharged structure (minor component)

dipolar ion, or zwitterion (major component)

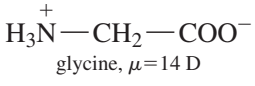
The dipolar nature of amino acids gives them some unusual properties:

1. Amino acids have **high melting points**, generally over 200 °C.

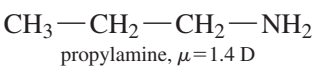


glycine, mp 262 °C

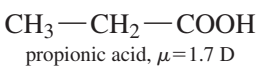
2. Amino acids are more **soluble in water** than they are in ether, dichloromethane, and other common organic solvents.
3. Amino acids have much **larger dipole moments** ($\mu$) than simple amines or simple acids.



glycine, $\mu = 14$ D



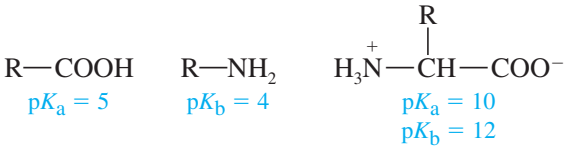
propylamine, $\mu = 1.4$ D



propionic acid, $\mu = 1.7$ D

4. Amino acids are **less acidic than most carboxylic acids** and **less basic than most amines**. In fact, the acidic part of the amino acid molecule is the $—NH_3^+$

group, not a —COOH group. The basic part is the $—COO^-$ group, and not a free $—NH_2$ group.




Because amino acids contain both acidic ($—NH_3^+$) and basic ($—COO^-$) groups, they are *amphoteric* (having both acidic and basic properties). The predominant form of the amino acid depends on the pH of the solution. In an acidic solution, the $—COO^-$ group is protonated to a free —COOH group, and the molecule has an overall positive charge. As the pH is raised, the —COOH loses its proton at about pH 2. This point is called $pK_{a1}$, the first acid-dissociation constant. As the pH is raised further, the $—NH_3^+$ group loses its proton at about pH 9 or 10. This point is called $pK_{a2}$, the second acid-dissociation constant. Above this pH, the molecule has an overall negative charge.

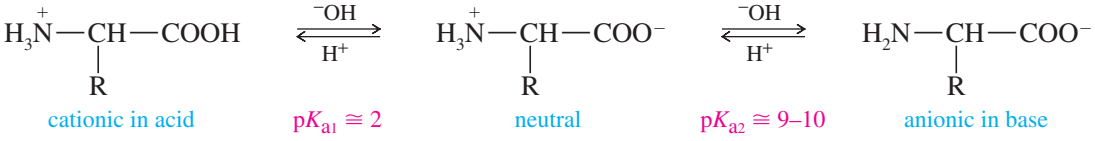



Figure 24-3 shows a titration curve for glycine. The curve starts at the bottom left, where glycine is entirely in its cationic form. Base is slowly added, and the pH is recorded. At pH 2.3, half of the cationic form has been converted to the zwitterionic form. At pH 6.0, essentially all the glycine is in the zwitterionic form. At pH 9.6, half of the zwitterionic form has been converted to the basic form. From this graph, we can see that glycine is mostly in the cationic form at pH values below 2.3, mostly in the zwitterionic form at pH values between 2.3 and 9.6, and mostly in the anionic form at pH values above 9.6. By varying the pH of the solution, we can control the charge on the molecule. This ability to control the charge of an amino acid is useful for separating and identifying amino acids by electrophoresis, as described in Section 24-4.

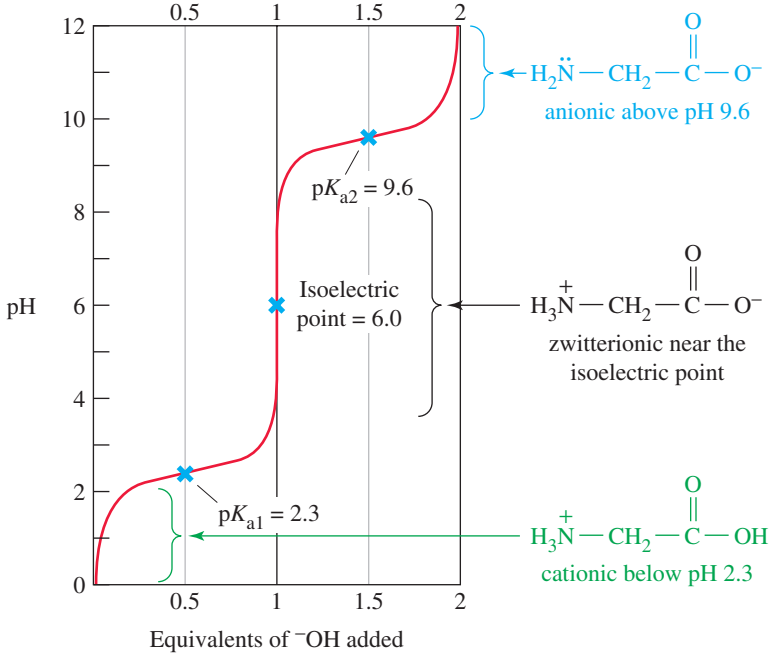
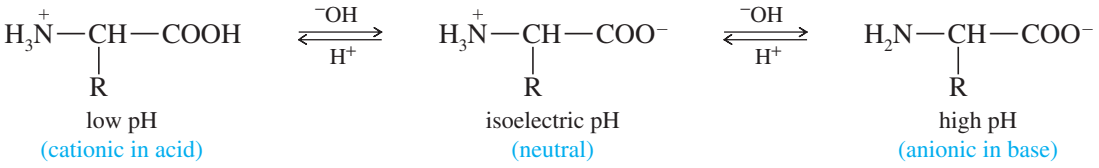



**FIGURE 24-3**
A titration curve for glycine. The pH controls the charge on glycine: cationic below pH 2.3; zwitterionic between pH 2.3 and 9.6; and anionic above pH 9.6. The isoelectric pH is 6.0.

## 24-4 Isoelectric Points and Electrophoresis

An amino acid bears a positive charge in acidic solution (low pH) and a negative charge in basic solution (high pH). There must be an intermediate pH where the amino acid is evenly balanced between the two forms, as the dipolar zwitterion with a net charge of zero. This pH is called the **isoelectric pH** or the **isoelectric point**.




The isoelectric points of the standard amino acids are given in Table 24-2. Notice that the isoelectric pH depends on the amino acid structure in a predictable way.

| | |
|---|---|
| acidic amino acids: | aspartic acid (2.8), glutamic acid (3.2) |
| neutral amino acids: | (5.0 to 6.3) |
| basic amino acids: | lysine (9.7), arginine (10.8), histidine (7.6) |

The side chains of aspartic acid and glutamic acid contain acidic carboxyl groups. These amino acids have acidic isoelectric points around pH 3. An acidic solution is needed to prevent deprotonation of the second carboxylic acid group and to keep the amino acid in its neutral isoelectric state.

Basic amino acids (histidine, lysine, and arginine) have isoelectric points at pH values of 7.6, 9.7, and 10.8, respectively. These values reflect the weak basicity of the imidazole ring, the intermediate basicity of an amino group, and the strong basicity of the guanidino group. A basic solution is needed in each case to prevent protonation of the basic side chain to keep the amino acid electrically neutral.

The other amino acids are considered neutral, with no strongly acidic or basic side chains. Their isoelectric points are slightly acidic (from about 5 to 6) because the $—NH_3^+$ group is slightly more acidic than the $—COO^-$ group is basic.

### PROBLEM 24-4

Draw the structure of the predominant form of

**(a)** isoleucine at pH 11 **(b)** proline at pH 2
**(c)** arginine at pH 7 **(d)** glutamic acid at pH 7
**(e)** a mixture of alanine, lysine, and aspartic acid at (i) pH 6; (ii) pH 11; (iii) pH 2

*problem-solving* **Hint**

At its isoelectric point (IEP), an amino acid has a net charge of zero, with $NH_3^+$ and $COO^-$ balancing each other. In more acidic solution (lower pH), the carboxyl group becomes protonated and the net charge is positive. In more basic solution (higher pH), the amino group loses its proton and the net charge is negative.

### PROBLEM 24-5

Draw the resonance forms of a protonated guanidino group, and explain why arginine has such a strongly basic isoelectric point.

### PROBLEM 24-6

Although tryptophan contains a heterocyclic amine, it is considered a neutral amino acid. Explain why the indole nitrogen of tryptophan is more weakly basic than one of the imidazole nitrogens of histidine.

**Electrophoresis** uses differences in isoelectric points to separate mixtures of amino acids (Figure 24-4). A streak of the amino acid mixture is placed in the center of a layer of acrylamide gel or a piece of filter paper wet with a buffer solution. Two electrodes are placed in contact with the edges of the gel or paper, and a potential of several thousand volts is applied across the electrodes. Positively charged (cationic) amino acids are attracted to the negative electrode (the cathode), and negatively charged (anionic) amino acids are attracted to the positive electrode (the anode). An amino acid at its isoelectric point has no net charge, so it does not move.

As an example, consider a mixture of alanine, lysine, and aspartic acid in a buffer solution at pH 6. Alanine is at its isoelectric point, in its dipolar zwitterionic form with

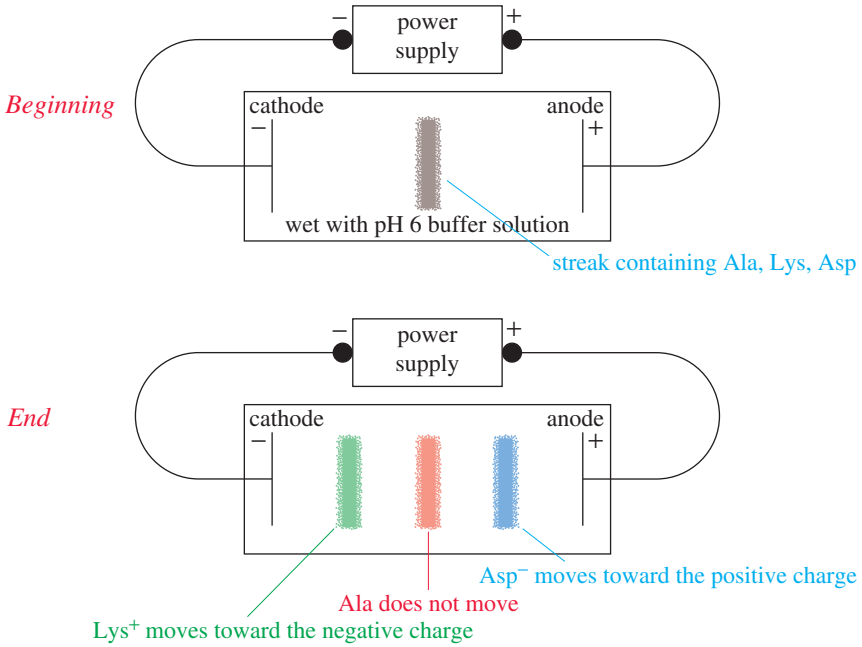
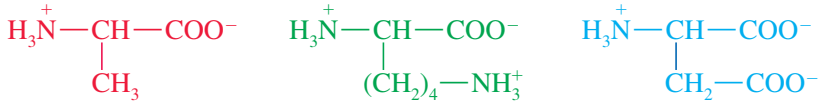



**FIGURE 24-4**
A simplified picture of the electrophoretic separation of alanine, lysine, and aspartic acid at pH 6. Cationic lysine is attracted to the cathode; anionic aspartic acid is attracted to the anode. Alanine is at its isoelectric point, so it does not move.

a net charge of zero. A pH of 6 is more acidic than the isoelectric pH for lysine (9.7), so lysine is in the cationic form. Aspartic acid has an isoelectric pH of 2.8, so it is in the anionic form.

*Structure at pH 6*




When a voltage is applied to a mixture of alanine, lysine, and aspartic acid at pH 6, alanine does not move. Lysine moves toward the negatively charged cathode, and aspartic acid moves toward the positively charged anode (Figure 24-4). After a period of time, the separated amino acids are recovered by cutting the paper or scraping the bands out of the gel. If electrophoresis is being used as an analytical technique (to determine the amino acids present in the mixture), the paper or gel is treated with a reagent such as ninhydrin (Section 24-9) to make the bands visible. Then the amino acids are identified by comparing their positions with those of standards.

**PROBLEM 24-7**

Draw the electrophoretic separation of Ala, Lys, and Asp at pH 9.7.

**PROBLEM 24-8**

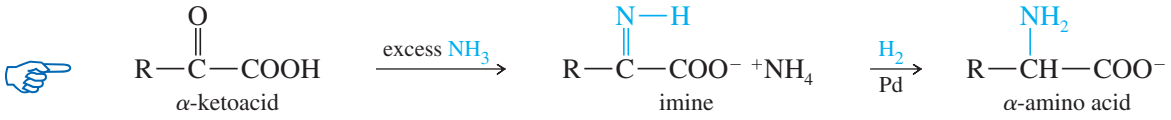
Draw the electrophoretic separation of Trp, Cys, and His at pH 6.0.

## 24-5 Synthesis of Amino Acids

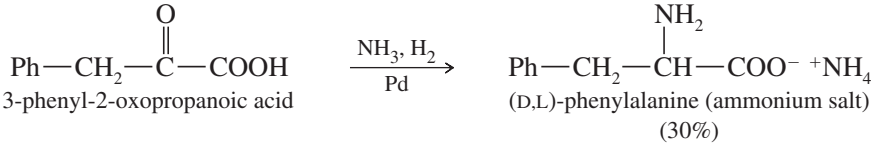
Naturally occurring amino acids can be obtained by hydrolyzing proteins and separating the amino acid mixture. Even so, it is often less expensive to synthesize the pure amino acid. In some cases, an unusual amino acid or an unnatural enantiomer is needed, and it must be synthesized. In this chapter, we consider four methods for making amino acids. All these methods are extensions of reactions we have already studied.

## 24-5A Reductive Amination

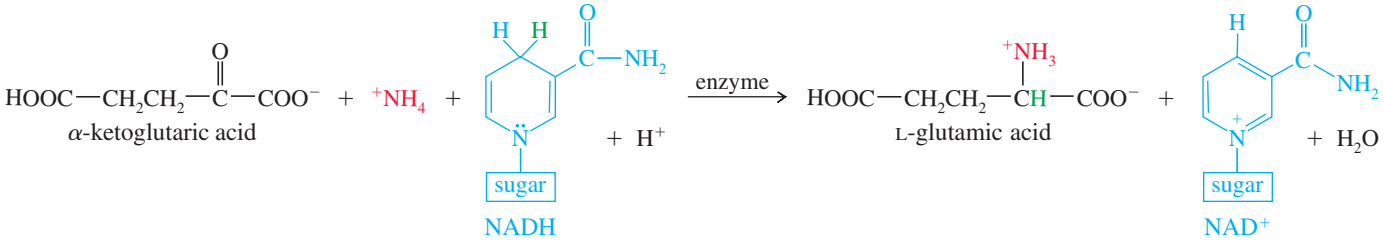
Reductive amination of ketones and aldehydes is one of the best methods for synthesizing amines (Section 19-19). It also forms amino acids. When an α-ketoacid is treated with ammonia, the ketone reacts to form an imine. The imine is reduced to an amine by hydrogen and a palladium catalyst. Under these conditions, the carboxylic acid is not reduced.



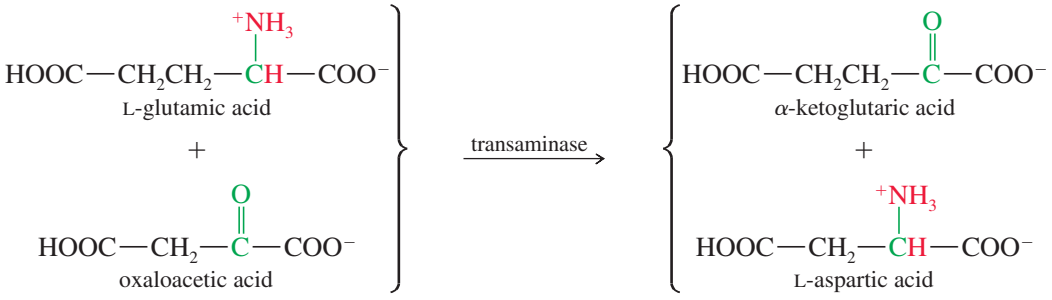

This entire synthesis is accomplished in one step by treating the α-ketoacid with ammonia and hydrogen in the presence of a palladium catalyst. The product is a racemic α-amino acid. The following reaction shows the synthesis of racemic phenylalanine from 3-phenyl-2-oxopropanoic acid.




We call reductive amination a **biomimetic** ("mimicking the biological process") synthesis because it resembles the biological synthesis of amino acids. The biosynthesis begins with reductive amination of α-ketoglutaric acid (an intermediate in the metabolism of carbohydrates), using ammonium ion as the aminating agent and NADH as the reducing agent. The product of this enzyme-catalyzed reaction is the pure L enantiomer of glutamic acid.




Biosynthesis of other amino acids uses L-glutamic acid as the source of the amino group. Such a reaction, moving an amino group from one molecule to another, is called a **transamination**, and the enzymes that catalyze these reactions are called *transaminases*. For example, the following reaction shows the biosynthesis of aspartic acid using glutamic acid as the nitrogen source. Once again, the enzyme-catalyzed biosynthesis gives the pure L enantiomer of the product.

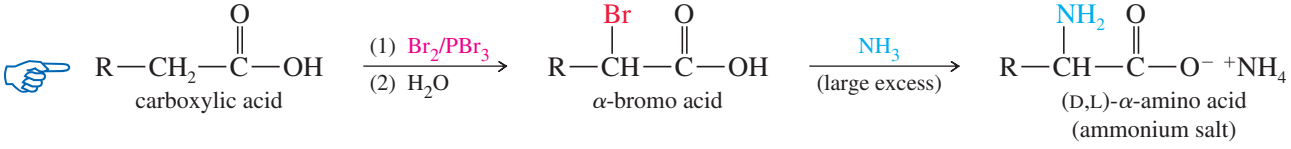


**PROBLEM 24-9**

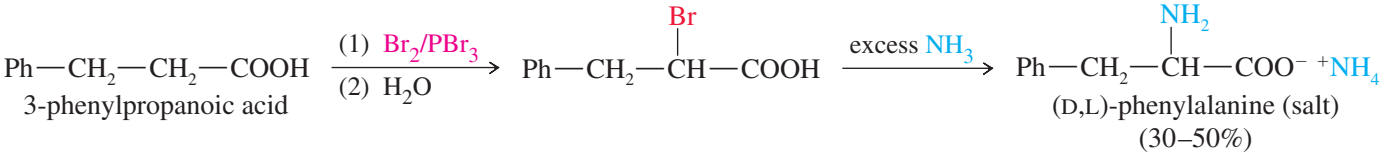
Show how the following amino acids might be formed in the laboratory by reductive amination of the appropriate α-ketoacid.
**(a)** alanine **(b)** leucine **(c)** serine **(d)** glutamine

## 24-5B Amination of an α- Halo Acid

The Hell–Volhard–Zelinsky reaction (Section 22-4) is an effective method for introducing bromine at the α position of a carboxylic acid. The racemic α-bromo acid is converted to a racemic α-amino acid by direct amination, using a large excess of ammonia.




In Section 19-19, we saw that direct alkylation is often a poor synthesis of amines, giving large amounts of overalkylated products. In this case, however, the reaction gives acceptable yields because a large excess of ammonia is used, making ammonia the nucleophile that is most likely to displace bromine. Also, the adjacent carboxylate ion in the product reduces the nucleophilicity of the amino group. The following sequence shows bromination of 3-phenylpropanoic acid, followed by displacement of bromide ion, to form the ammonium salt of racemic phenylalanine.

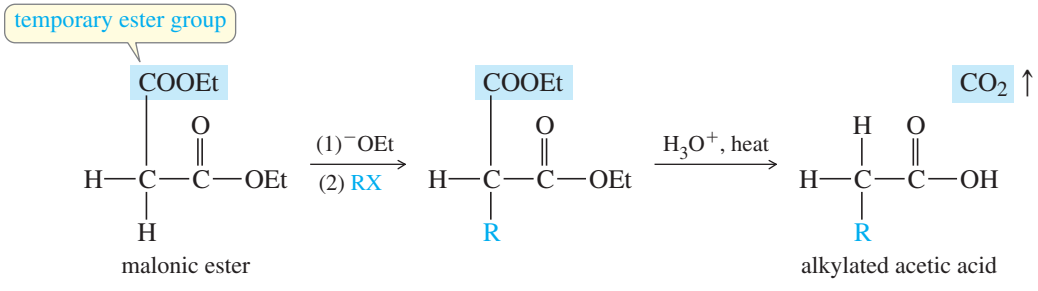



**PROBLEM 24-10**

Show how you would use bromination followed by amination to synthesize the following amino acids.
**(a)** glycine **(b)** leucine **(c)** glutamic acid

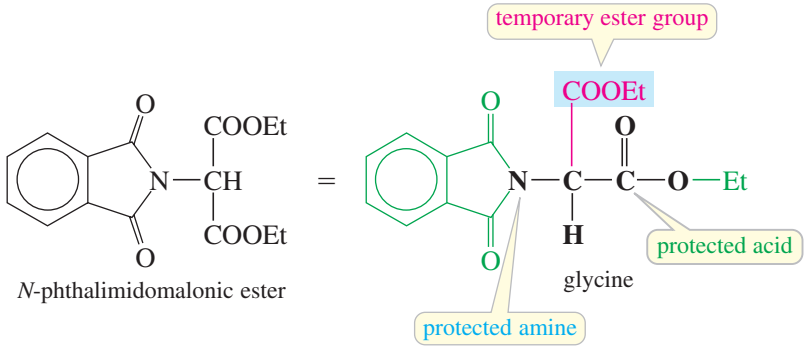
## 24-5C The Gabriel–Malonic Ester Synthesis

One of the best methods of amino acid synthesis is a combination of the Gabriel synthesis of amines (Section 19-21) with the malonic ester synthesis of carboxylic acids (Section 22-16). The conventional malonic ester synthesis involves alkylation of diethyl malonate, followed by hydrolysis and decarboxylation to give an alkylated acetic acid.



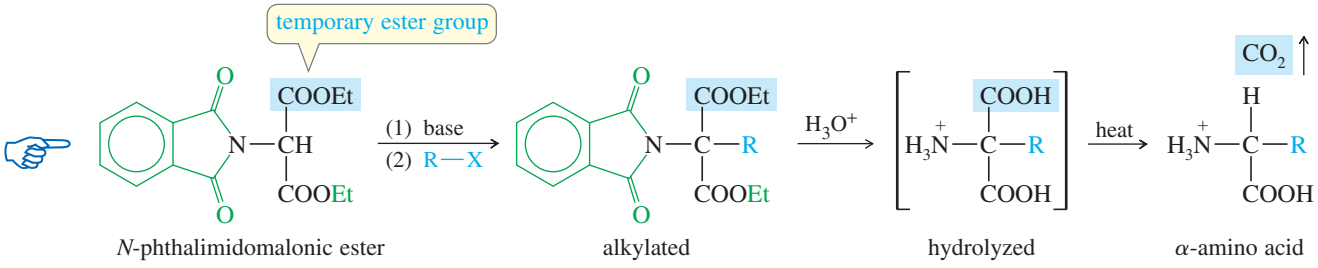

To adapt this synthesis to making amino acids, we begin with a malonic ester that contains an α-amino group. The amino group is protected as a non-nucleophilic amide to prevent it from attacking the alkylating agent (RX).

The Gabriel–malonic ester synthesis begins with *N*-phthalimidomalonic ester. Think of *N*-phthalimidomalonic ester as a molecule of glycine (aminoacetic acid) with the amino group protected as an amide (a phthalimide in this case) to keep it from acting as a nucleophile. The acid is protected as an ethyl ester, and the $\alpha$ position is further activated by the additional (temporary) ester group of diethyl malonate.

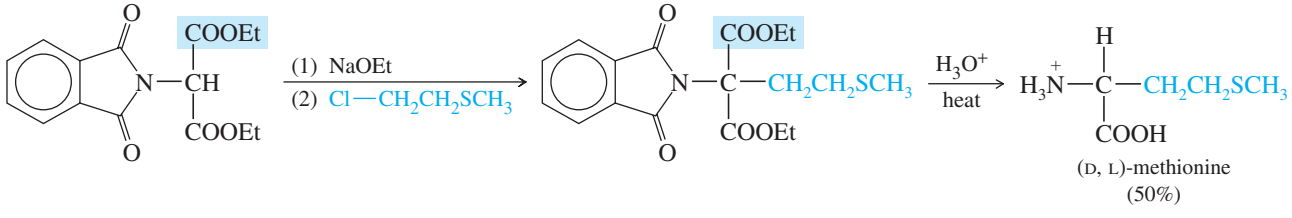



Just as the malonic ester synthesis gives substituted acetic acids, the *N*-phthalimidomalonic ester synthesis gives substituted aminoacetic acids: $\alpha$-amino acids. *N*-Phthalimidomalonic ester is alkylated in the same way as malonic ester. When the alkylated *N*-phthalimidomalonic ester is hydrolyzed, the phthalimido group is hydrolyzed along with the ester groups. The product is an alkylated aminomalonic acid. Decarboxylation gives a racemic $\alpha$-amino acid.

*The Gabriel–malonic ester synthesis*




The Gabriel–malonic ester synthesis is used to make many amino acids that cannot be formed by direct amination of haloacids. The following example shows the synthesis of methionine, which is formed in very poor yield by direct amination.

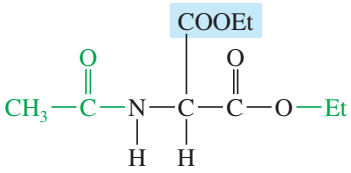



**PROBLEM 24-11**

Show how the Gabriel–malonic ester synthesis could be used to make
**(a)** valine **(b)** phenylalanine **(c)** glutamic acid **(d)** leucine

***PROBLEM 24-12**

The Gabriel–malonic ester synthesis uses an aminomalonic ester with the amino group protected as a phthalimide. A variation has the amino group protected as an acetamido group. Propose how you might use an *acetamidomalonic ester* synthesis to make phenylalanine.

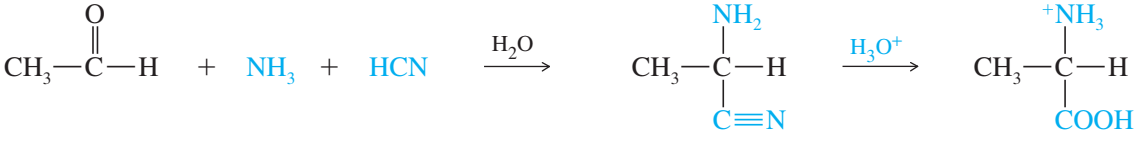


acetamidomalonic ester

## 24-5D The Strecker Synthesis

The first known synthesis of an amino acid occurred in 1850 in the laboratory of Adolph Strecker in Tübingen, Germany. Strecker added acetaldehyde to an aqueous solution of ammonia and HCN. The product was α-amino propionitrile, which Strecker hydrolyzed to racemic alanine.

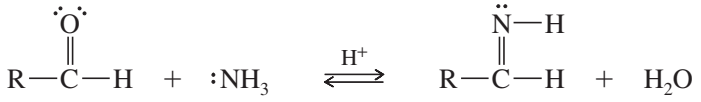
*The Strecker synthesis of alanine*



acetaldehyde — α-amino propionitrile — (D,L)-alanine (60%)

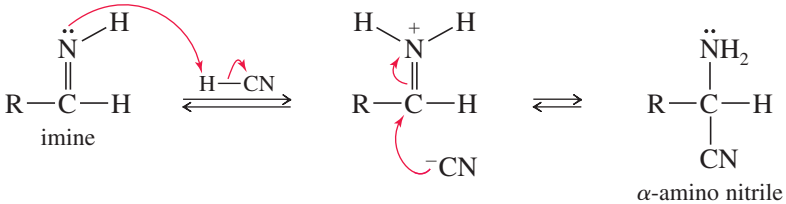
The **Strecker synthesis** can form a large number of amino acids from appropriate aldehydes. The mechanism is shown next. First, the aldehyde reacts with ammonia to give an imine. The imine is a nitrogen analogue of a carbonyl group, and it is electrophilic when protonated. Attack of cyanide ion on the protonated imine gives the α-amino nitrile. This mechanism is similar to that for formation of a cyanohydrin (Section 18-15), except that in the Strecker synthesis cyanide ion attacks an imine rather than the aldehyde itself.

*Step 1: The aldehyde reacts with ammonia to form the imine* (mechanism in Section 18-16)



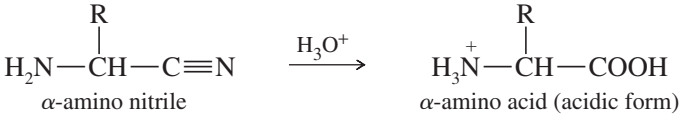
aldehyde — imine

*Step 2: Cyanide ion attacks the imine.*



imine — α-amino nitrile

In a separate step, hydrolysis of the α-amino nitrile (Section 21-7D) gives an α-amino acid.



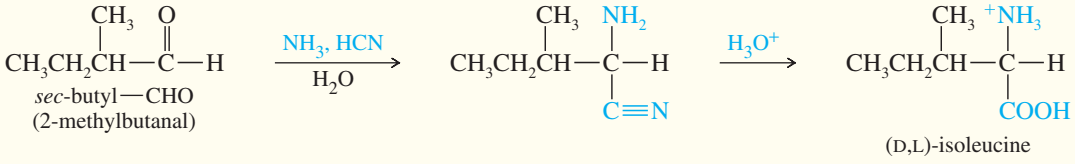
α-amino nitrile — α-amino acid (acidic form)

## SOLVED PROBLEM 24-1

Show how you would use a Strecker synthesis to make isoleucine.

## SOLUTION

Isoleucine has a *sec*-butyl group for its side chain. Remember that $CH_3—CHO$ undergoes Strecker synthesis to give alanine, with $CH_3$ as the side chain. Therefore, *sec*-butyl—CHO should give isoleucine.




## PROBLEM 24-13

**(a)** Show how you would use a Strecker synthesis to make phenylalanine.
**(b)** Propose a mechanism for each step in the synthesis in part (a).

## PROBLEM 24-14

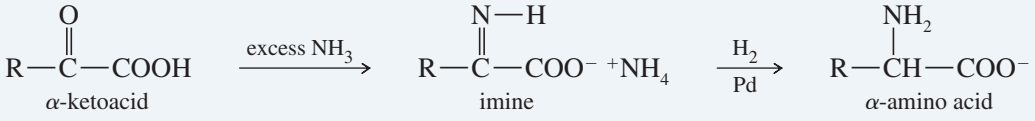
Show how you would use a Strecker synthesis to make
**(a)** leucine **(b)** valine **(c)** aspartic acid

**problem-solving Hint**

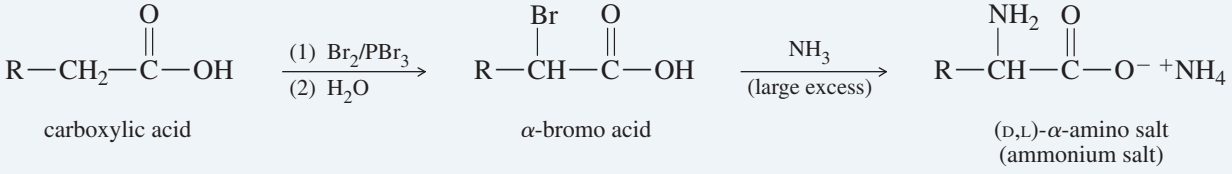
In the malonic ester synthesis, use the side chain of the desired amino acid (must be a good $S_N2$ substrate) to alkylate the ester. In the Strecker synthesis, the aldehyde carbon becomes the $\alpha$ carbon of the amino acid: Begin with [side chain]—CHO.

## SUMMARY Syntheses of Amino Acids

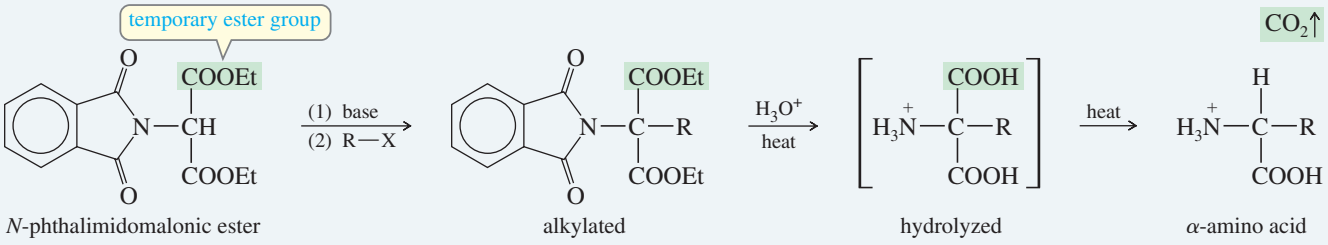
**1.** *Reductive amination* (Section 24-5A)



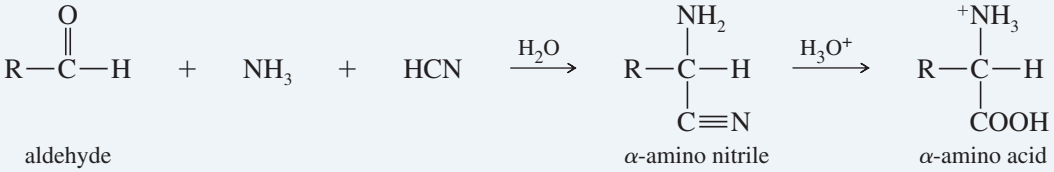

**2.** *Amination of an α-haloacid* (Section 24-5B)




**3.** *The Gabriel–malonic ester synthesis* (Section 24-5C)




**4.** *The Strecker synthesis* (Section 24-5D)



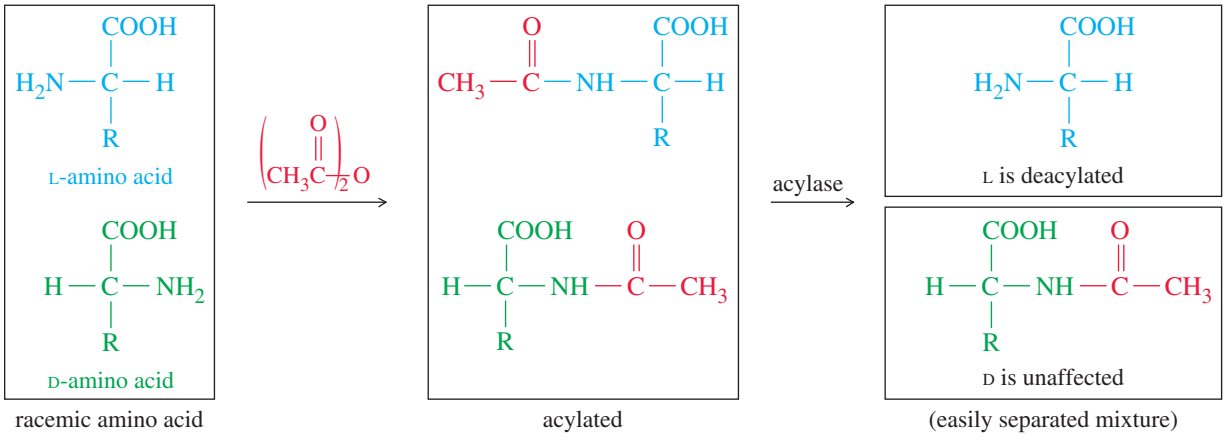



**FIGURE 24-5**
Selective enzymatic deacylation. An acylase enzyme (such as hog kidney acylase or carboxypeptidase) deacylates only the natural L-amino acid.

## 24-6 Resolution of Amino Acids

All the laboratory syntheses of amino acids described in Section 24-5 produce racemic products. In most cases, only the L enantiomers are biologically active. The D enantiomers may even be toxic. Pure L enantiomers are needed for peptide synthesis if the product is to have the activity of the natural material. Therefore, we must be able to resolve a racemic amino acid into its enantiomers.

In many cases, amino acids can be resolved by the methods we have already discussed (Section 5-16). If a racemic amino acid is converted to a salt with an optically pure chiral acid or base, two diastereomeric salts are formed. These salts can be separated by physical means such as selective crystallization or chromatography. Pure enantiomers are then regenerated from the separated diastereomeric salts. Strychnine and brucine are naturally occurring optically active bases, and tartaric acid is used as an optically active acid for resolving racemic mixtures.

**Enzymatic resolution** is also used to separate the enantiomers of amino acids. Enzymes are chiral molecules with specific catalytic activities. For example, when an acylated amino acid is treated with an enzyme like hog kidney acylase or carboxypeptidase, the enzyme cleaves the acyl group from just the molecules having the natural (L) configuration. The enzyme does not recognize D-amino acids, so they are unaffected. The resulting mixture of acylated D-amino acid and deacylated L-amino acid is easily separated. Figure 24-5 shows how this selective enzymatic deacylation is accomplished.

**PROBLEM 24-15**

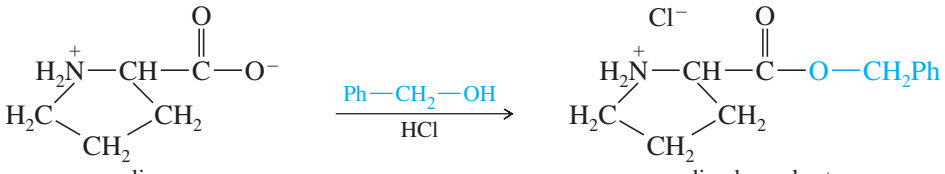
Suggest how you would separate the free L-amino acid from its acylated D enantiomer in Figure 24-5.

## 24-7 Reactions of Amino Acids

Amino acids undergo many of the standard reactions of both amines and carboxylic acids. Conditions for some of these reactions must be carefully selected, however, so that the amino group does not interfere with a carboxyl group reaction, and vice versa. We will consider two of the most useful reactions, esterification of the carboxyl group and acylation of the amino group. These reactions are often used to protect either the carboxyl group or the amino group while the other group is being modified or coupled to another amino acid. Amino acids also undergo reactions that are specific to the $\alpha$-amino acid structure. One of these unique amino acid reactions is the formation of a colored product on treatment with ninhydrin, discussed in Section 24-7C.

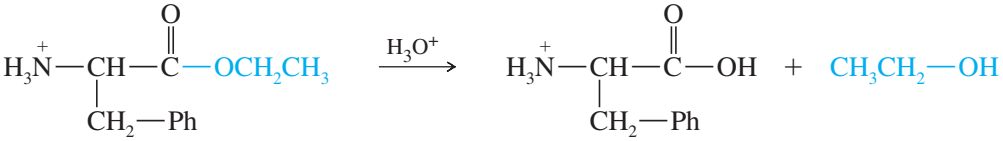
## 24-7A Esterification of the Carboxyl Group

Like monofunctional carboxylic acids, amino acids are esterified by treatment with a large excess of an alcohol and an acidic catalyst (often gaseous HCl). Under these acidic conditions, the amino group is present in its protonated ($-NH_3^+$) form, so it does not interfere with esterification. The following example illustrates esterification of an amino acid.



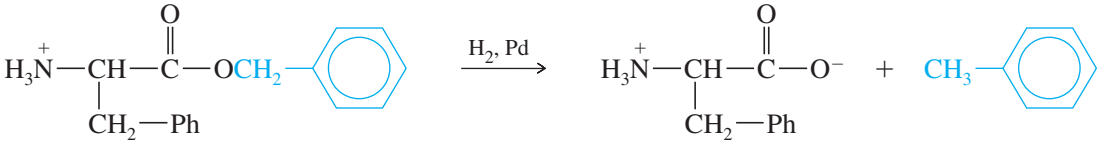

proline → proline benzyl ester (90%)

Esters of amino acids are often used as protected derivatives to prevent the carboxyl group from reacting in some undesired manner. Methyl, ethyl, and benzyl esters are the most common protecting groups. Aqueous acid hydrolyzes the ester and regenerates the free amino acid.




phenylalanine ethyl ester → phenylalanine + $CH_3CH_2—OH$

Benzyl esters are particularly useful as protecting groups because they can be removed either by acidic hydrolysis or by neutral **hydrogenolysis** ("breaking apart by addition of hydrogen"). Catalytic hydrogenation cleaves the benzyl ester, converting the benzyl group to toluene and leaving the deprotected amino acid. Although the mechanism of this hydrogenolysis is not well known, it apparently hinges on the ease of formation of benzylic intermediates.




phenylalanine benzyl ester → phenylalanine + toluene

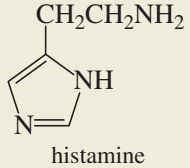
**PROBLEM 24-16**

Propose a mechanism for the acid-catalyzed hydrolysis of phenylalanine ethyl ester.

**PROBLEM 24-17**

Give equations for the formation and hydrogenolysis of glutamine benzyl ester.

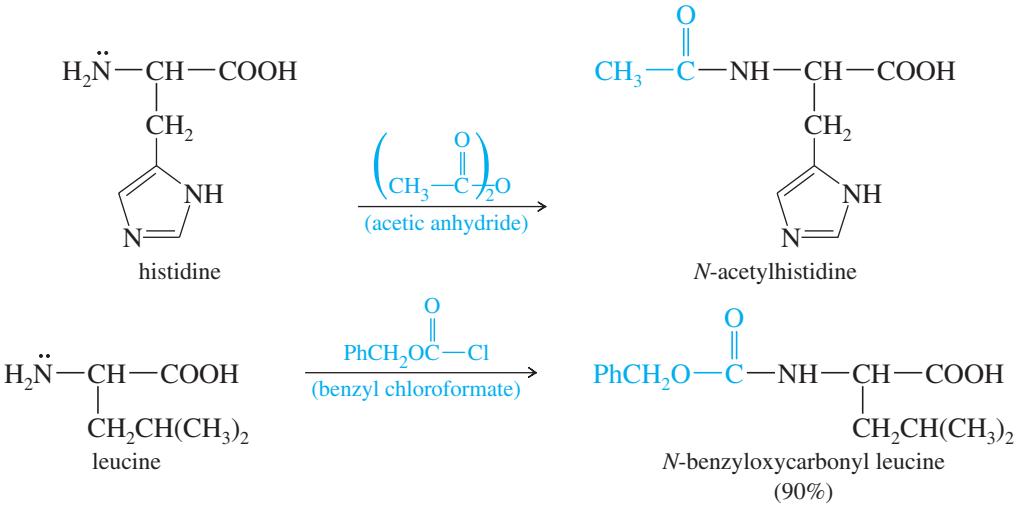
Decarboxylation is an important reaction of amino acids in many biological processes. Histamine, which causes runny noses and itchy eyes, is synthesized in the body by decarboxylation of histidine. The enzyme that catalyzes this reaction is called histidine decarboxylase.



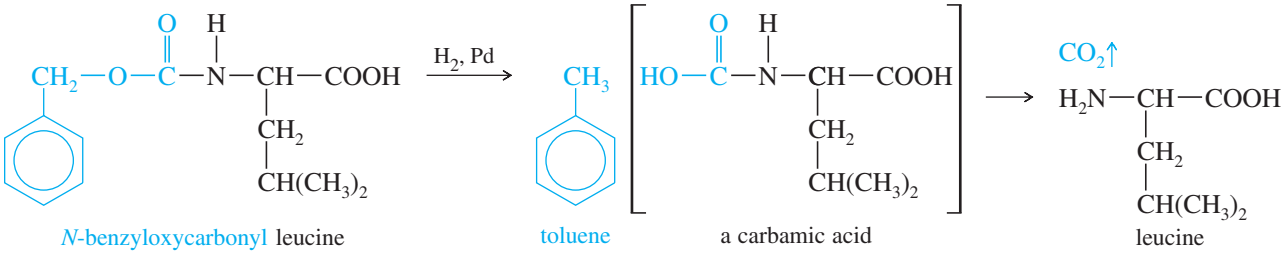
histamine

## 24-7B Acylation of the Amino Group: Formation of Amides

Just as an alcohol esterifies the carboxyl group of an amino acid, an acylating agent converts the amino group to an amide. Acylation of the amino group is often done to protect it from unwanted nucleophilic reactions. A wide variety of acid chlorides and anhydrides are used for acylation. Benzyl chloroformate acylates the amino group to give a benzyloxycarbonyl derivative, often used as a protecting group in peptide synthesis (Section 24-10).




The amino group of the *N*-benzyloxycarbonyl derivative is protected as the amide half of a carbamate ester (a urethane, Section 21-16), which is more easily hydrolyzed than most other amides. In addition, the ester half of this urethane is a benzyl ester that undergoes hydrogenolysis. Catalytic hydrogenolysis of the *N*-benzyloxycarbonyl amino acid gives an unstable carbamic acid that quickly decarboxylates to give the deprotected amino acid.



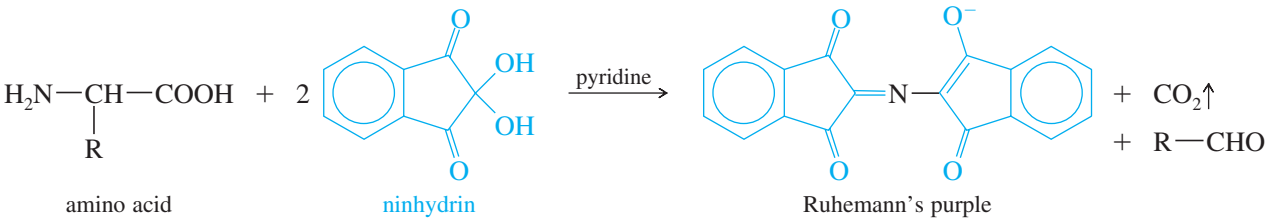

**PROBLEM 24-18**

Give equations for the formation and hydrogenolysis of *N*-benzyloxycarbonyl methionine.

## 24-7C Reaction with Ninhydrin

*Ninhydrin* is a common reagent for visualizing spots or bands of amino acids that have been separated by chromatography or electrophoresis. When ninhydrin reacts with an amino acid, one of the products is a deep violet, resonance-stabilized anion called *Ruhemann's purple*. Ninhydrin produces this same purple dye regardless of the structure of the original amino acid. The side chain of the amino acid is lost as an aldehyde.

Reaction of an amino acid with ninhydrin




The reaction of amino acids with ninhydrin can detect amino acids on a wide variety of substrates. For example, if a kidnapper touches a ransom note with his fingers, the dermal ridges on his fingers leave traces of amino acids from skin secretions.

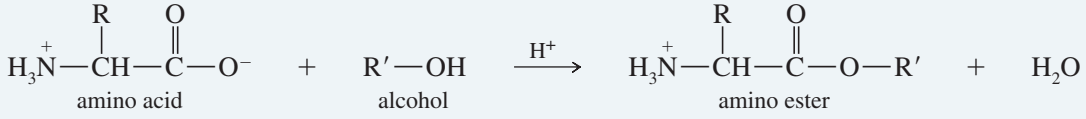
Treatment of the paper with ninhydrin and pyridine causes these secretions to turn purple, forming a visible fingerprint.

**PROBLEM 24-19**

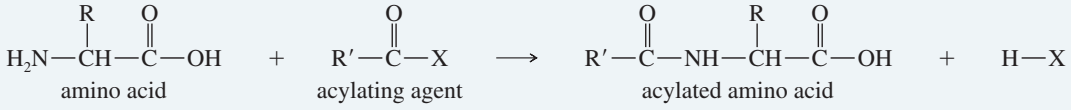
Use resonance forms to show delocalization of the negative charge in the Ruhemann's purple anion.

## SUMMARY Reactions of Amino Acids

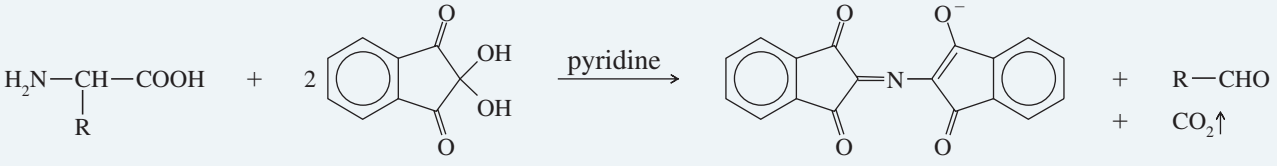
**1.** *Esterification of the carboxyl group* (Section 24-7A)



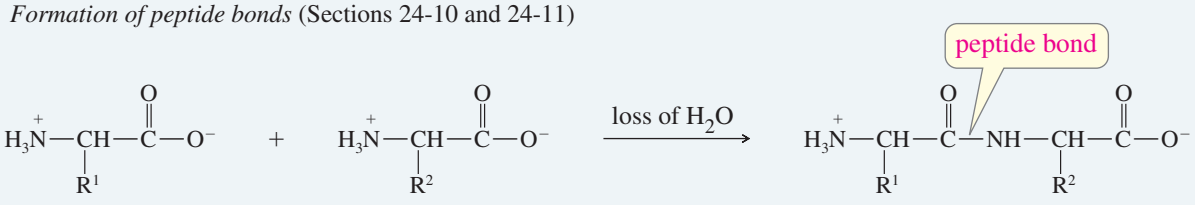

**2.** *Acylation of the amino group: formation of amides* (Section 24-7B)




**3.** *Reaction with ninhydrin* (Section 24-7C)




**4.** *Formation of peptide bonds* (Sections 24-10 and 24-11)

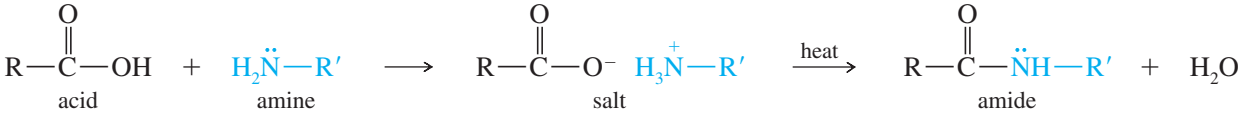



Amino acids also undergo many other common reactions of amines and acids.

## 24-8 Structure and Nomenclature of Peptides and Proteins

### 24-8A Peptide Structure

The most important reaction of amino acids is the formation of peptide bonds. Amines and acids can condense, with the loss of water, to form amides. Industrial processes often make amides simply by mixing the acid and the amine, then heating the mixture to drive off water.




Recall from Section 21-13 that amides are the most stable acid derivatives. This stability is partly due to the strong resonance interaction between the nonbonding electrons on nitrogen and the carbonyl group. The amide nitrogen is no longer a strong base, and the C—N bond has restricted rotation because of its partial double-bond character. Figure 24-6 shows the resonance forms we use to explain the partial double-bond

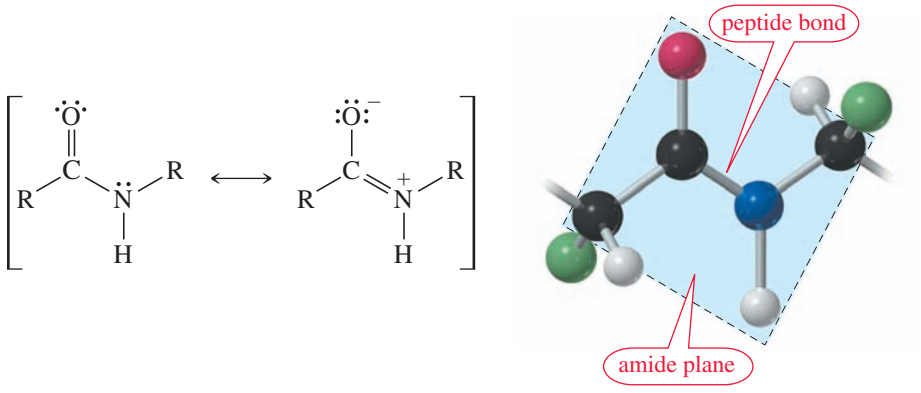
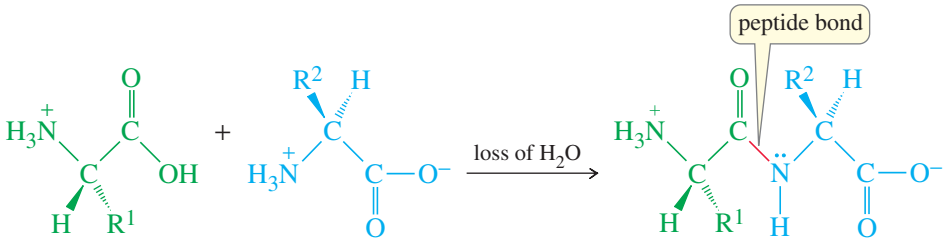



**FIGURE 24-6**
Resonance stabilization of an amide accounts for its enhanced stability, the weak basicity of the nitrogen atom, and the restricted rotation of the C—N bond. In a peptide, the amide bond is called a peptide bond. It holds six atoms in a plane: the C and O of the carbonyl, the N and its H, and the two associated $\alpha$ carbon atoms.

character and restricted rotation of an amide bond. In a peptide, this partial double-bond character results in six atoms being held rather rigidly in a plane.

Having both an amino group and a carboxyl group, an amino acid is ideally suited to form an amide linkage. Under the proper conditions, the amino group of one molecule condenses with the carboxyl group of another. The product is an amide called a *dipeptide* because it consists of two amino acids. The amide linkage between the amino acids is called a **peptide bond**. Although it has a special name, a peptide bond is just like other amide bonds we have studied.




In this manner, any number of amino acids can be bonded in a continuous chain. A **peptide** is a compound containing two or more amino acids linked by amide bonds between the amino group of each amino acid and the carboxyl group of the neighboring amino acid. Each amino acid unit in the peptide is called a **residue**. A **polypeptide** is a peptide containing many amino acid residues but usually having a molecular weight of less than about 5000. **Proteins** contain more amino acid units, with molecular weights ranging from about 6000 to about 40,000,000. The term **oligopeptide** is occasionally used for peptides containing about four to ten amino acid residues. Figure 24-7 shows the structure of the nonapeptide bradykinin, a human hormone that helps to control blood pressure.

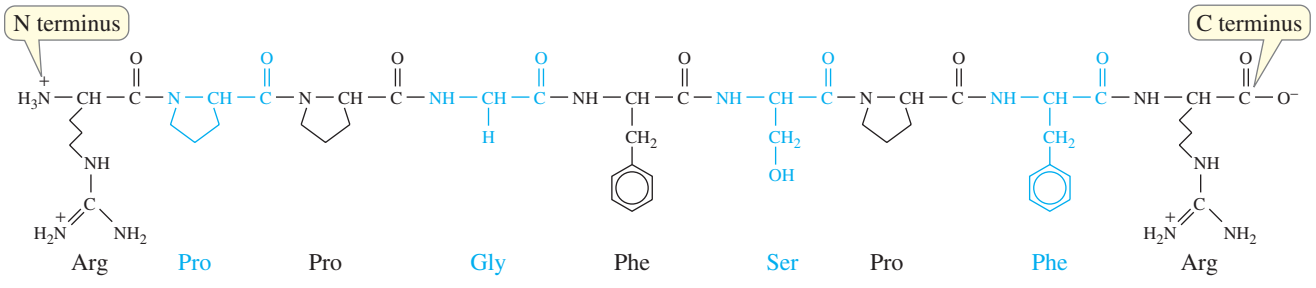
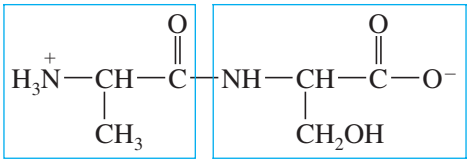



**FIGURE 24-7**
The human hormone bradykinin is a nonapeptide with a free $-NH_3^+$ at its N terminus and a free $-COO^-$ at its C terminus.

The end of the peptide with the free amino group ($-NH_3^+$) is called the **N-terminal end** or the **N terminus**, and the end with the free carboxyl group ($-COO^-$) is called the **C-terminal end** or the **C terminus**. Peptide structures are generally drawn with the N terminus at the left and the C terminus at the right, as bradykinin is drawn in Figure 24-7.

## 24-8B Peptide Nomenclature

The names of peptides reflect the names of the amino acid residues involved in the amide linkages, beginning at the *N* terminus. All except the last are given the *-yl* suffix of acyl groups. For example, the following dipeptide is named alanylserine. The alanine residue has the *-yl* suffix because it has acylated the nitrogen of serine.




Ala-Ser

Bradykinin (Figure 24-7) is named as follows (without any spaces):

arginyl prolyl prolyl glycyl phenylalanyl seryl prolyl phenylalanyl arginine

This is a cumbersome and awkward name. A shorthand system is more convenient, representing each amino acid by its three-letter abbreviation. These abbreviations, given in Table 24-2, are generally the first three letters of the name. Once again, the amino acids are arranged from the N terminus at the left to the C terminus at the right. Bradykinin has the following abbreviated name:

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

Single-letter symbols (also given in Table 24-2) are becoming widely used as well. Using single letters, we symbolize bradykinin by

RPPGFSPFR

**PROBLEM 24-20**

Draw the complete structures of the following peptides:
**(a)** Thr-Phe-Met **(b)** serylarginylglycylphenylalanine **(c)** IMQDK **(d)** ELVIS

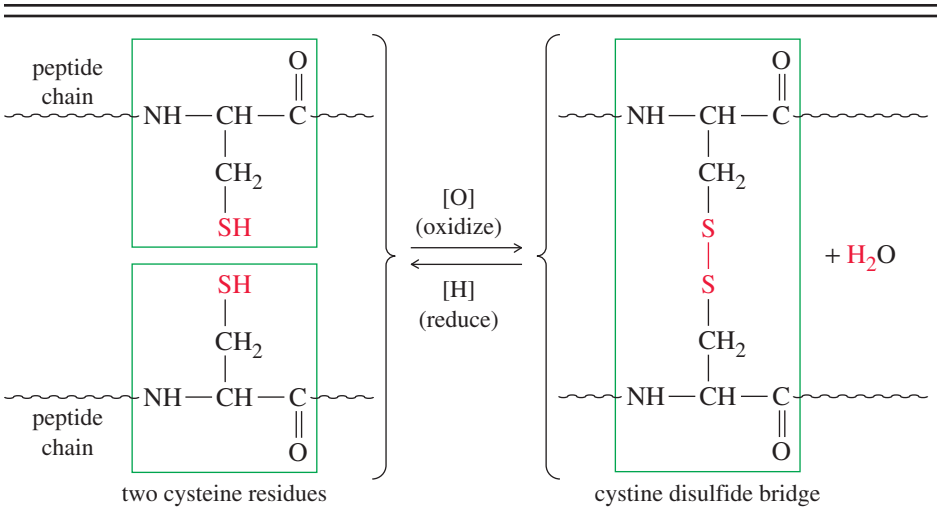
## 24-8C Disulfide Linkages

Amide linkages (peptide bonds) form the backbone of the amino acid chains we call peptides and proteins. A second kind of covalent bond is possible between any cysteine residues present. Cysteine residues can form **disulfide bridges** (also called **disulfide linkages**) which can join two chains or link a single chain into a ring.

Mild oxidation joins two molecules of a thiol into a disulfide, forming a disulfide linkage between the two thiol molecules. This reaction is reversible, and a mild reduction cleaves the disulfide.

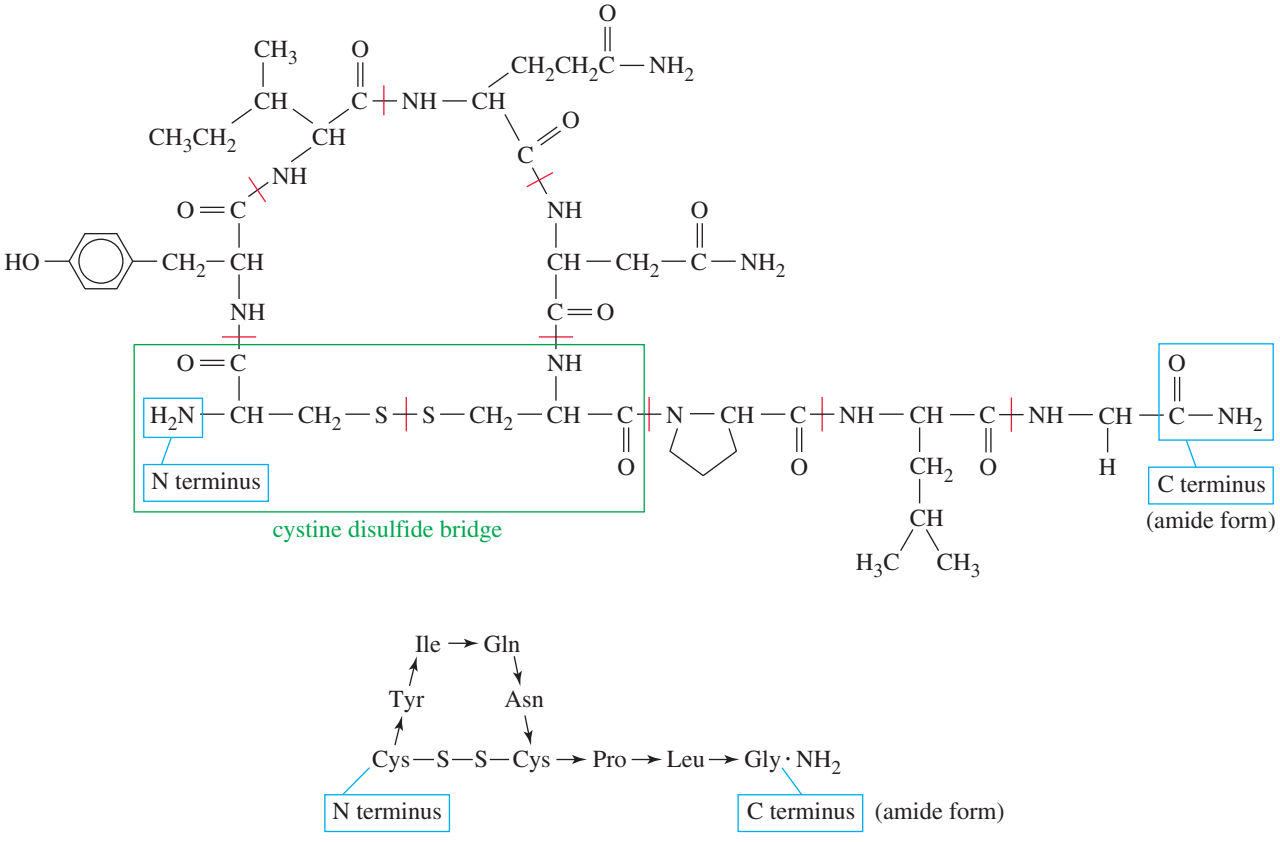
$$\underset{\text{two molecules of thiol}}{R-SH + HS-R} \underset{\text{[reduction]}}{\overset{\text{[oxidation]}}{\rightleftharpoons}} \underset{\text{disulfide}}{R-S-S-R} + H_2O$$

Similarly, two cysteine sulfhydryl ($-SH$) groups are oxidized to give a disulfide-linked pair of amino acids. This disulfide-linked dimer of cysteine is called *cystine*. Figure 24-8 shows formation of a cystine disulfide bridge linking two peptide chains.




■ **FIGURE 24-8**
Cystine, a dimer of cysteine, results when two cysteine residues are oxidized to form a disulfide bridge.

Two cysteine residues may form a disulfide bridge within a single peptide chain, making a ring. Figure 24-9 shows the structure of human oxytocin, a peptide hormone that causes contraction of uterine smooth muscle and induces labor. Oxytocin is a nonapeptide with two cysteine residues (at positions 1 and 6) linking part of the molecule in a large ring. In drawing the structure of a complicated peptide, arrows are often used to connect the amino acids, showing the direction from N terminus to C terminus. Notice that the C terminus of oxytocin is a primary amide $(\text{Gly} \cdot \text{NH}_2)$ rather than a free carboxyl group.




■ **FIGURE 24-9**
Structure of human oxytocin. A disulfide linkage holds part of the molecule in a large ring.

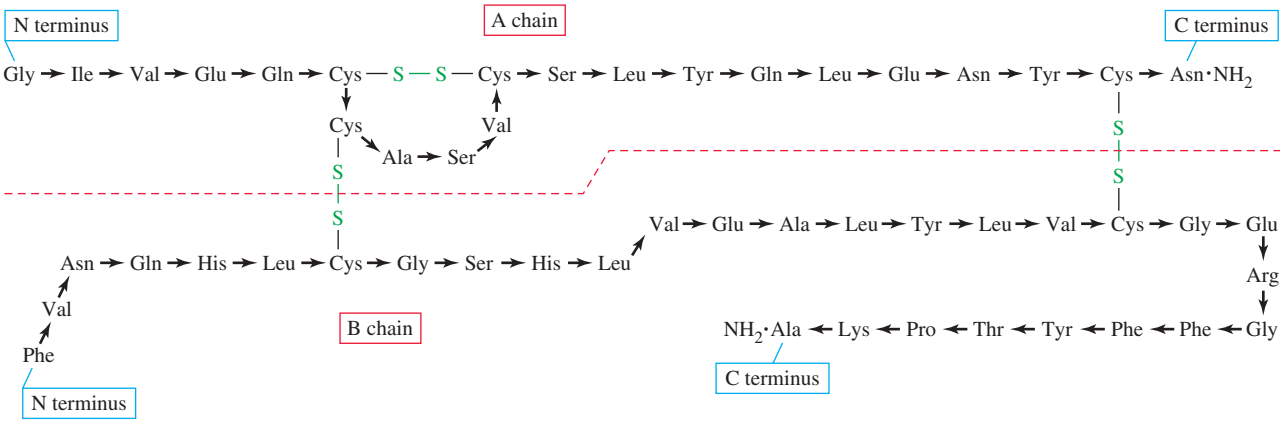



**FIGURE 24-10**
Structure of insulin. Two chains are joined at two positions by disulfide bridges, and a third disulfide bond holds the A chain in a ring.

Orexin A (from the Greek ***orexis***, "appetite") is a 33 amino acid neuropeptide connected by two disulfide bridges. Orexin A is a powerful stimulant for food intake and gastric juice secretion. Scientists are studying orexin A to learn more about the regulation of appetite and eating, hoping to learn more about causes and potential treatments for anorexia nervosa.

Figure 24-10 shows the structure of insulin, a more complex peptide hormone that regulates glucose metabolism. Insulin is composed of two separate peptide chains, the *A chain*, containing 21 amino acid residues, and the *B chain*, containing 30. The A and B chains are joined at two positions by disulfide bridges, and the A chain has an additional disulfide bond that holds six amino acid residues in a ring. The C-terminal amino acids of both chains occur as primary amides.

Disulfide bridges are commonly manipulated in the process of giving hair a *permanent wave*. Hair is composed of protein, which is made rigid and tough partly by disulfide bonds. When hair is treated with a solution of a thiol such as 2-mercaptoethanol ($HS—CH_2—CH_2—OH$), the disulfide bridges are reduced and cleaved. The hair is wrapped around curlers, and the disulfide bonds are allowed to re-form, either by air oxidation or by application of a *neutralizer*. The disulfide bonds re-form in new positions, holding the hair in the bent conformation enforced by the curlers.

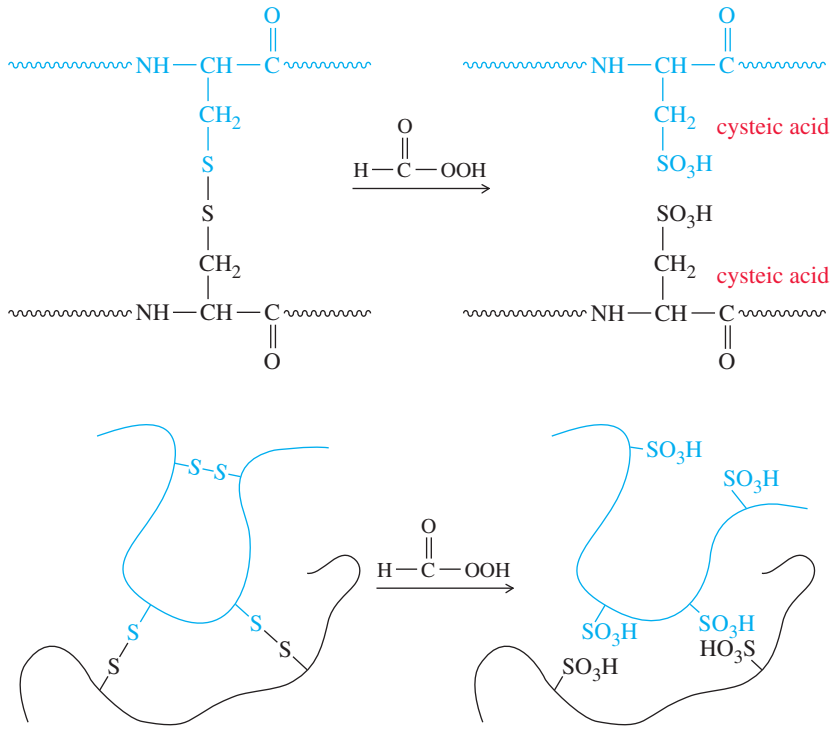
## 24-9 Peptide Structure Determination

Insulin is a relatively simple protein, yet it is a complicated organic structure. How is it possible to determine the complete structure of a protein with hundreds of amino acid residues and a molecular weight of many thousands? Chemists have developed clever ways to determine the exact sequence of amino acids in a protein. We will consider some of the most common methods.

### 24-9A Cleavage of Disulfide Linkages

The first step in structure determination is to break all the disulfide bonds, opening any disulfide-linked rings and separating the individual peptide chains. The individual peptide chains are then purified and analyzed separately.

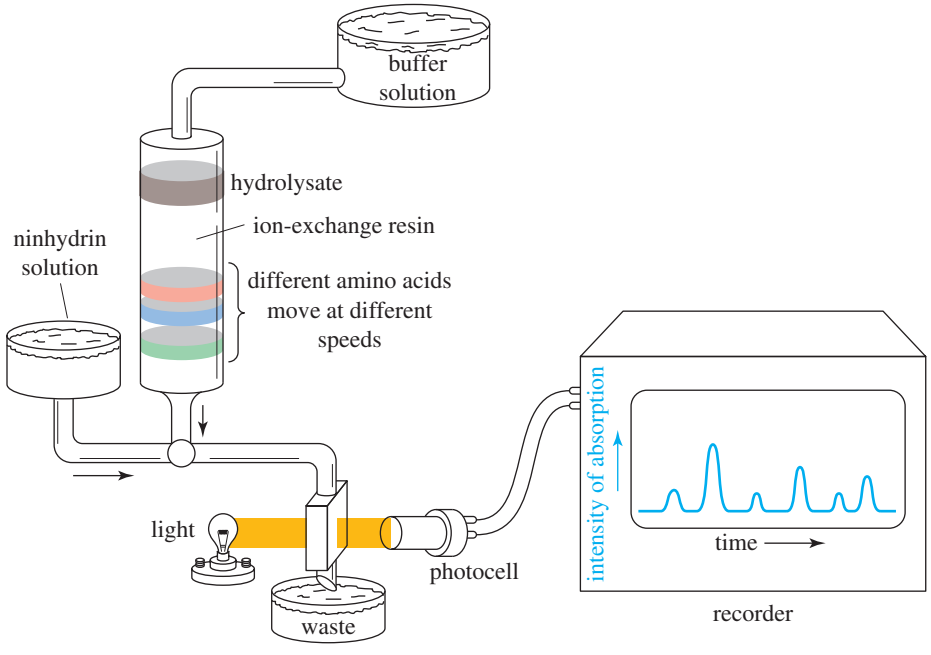
Cystine bridges are easily cleaved by reducing them to the thiol (cysteine) form. These reduced cysteine residues have a tendency to reoxidize and re-form disulfide bridges, however. A more permanent cleavage involves oxidizing the disulfide linkages with peroxyformic acid (Figure 24-11). This oxidation converts the disulfide bridges to sulfonic acid ($—SO_3H$) groups. The oxidized cysteine units are called *cysteic acid* residues.




■ **FIGURE 24-11**
Oxidation of a protein by peroxyformic acid cleaves all the disulfide linkages by oxidizing cystine to cysteic acid.

## 24-9B Determination of the Amino Acid Composition

Once the disulfide bridges have been broken and the individual peptide chains have been separated and purified, the structure of each chain must be determined. The first step is to determine which amino acids are present and in what proportions. To analyze the amino acid composition, the peptide chain is completely hydrolyzed by boiling it for 24 hours in 6 M HCl. The resulting mixture of amino acids (the *hydrolysate*) is placed on the column of an *amino acid analyzer*, diagrammed in Figure 24-12.




■ **FIGURE 24-12**
In an amino acid analyzer, the hydrolysate passes through an ion-exchange column. The solution emerging from the column is treated with ninhydrin, and its absorbance is recorded as a function of time. Each amino acid is identified by the retention time required to pass through the column.

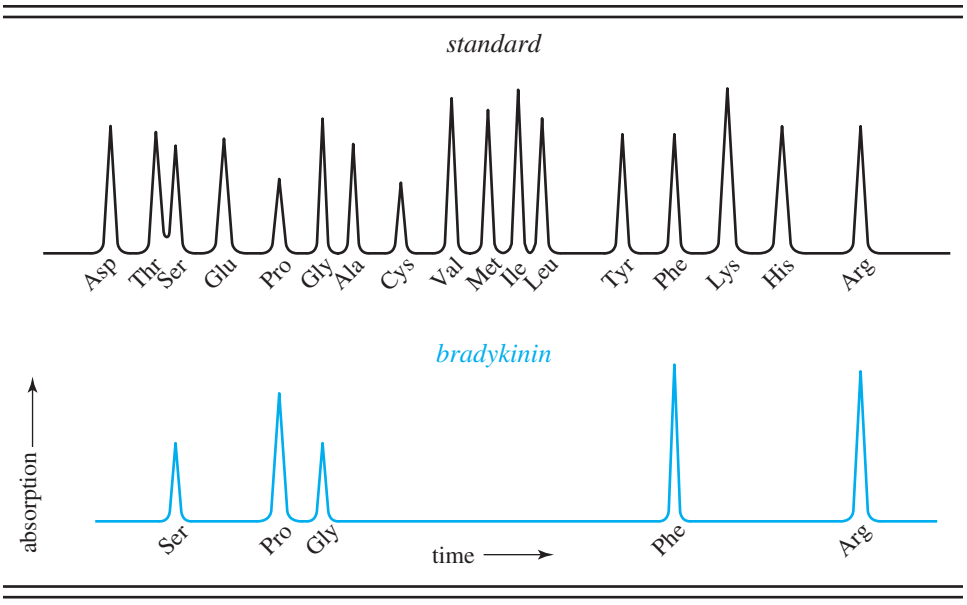



**FIGURE 24-13**
Use of an amino acid analyzer to determine the composition of human bradykinin. The bradykinin peaks for Pro, Arg, and Phe are larger than those in the standard equimolar mixture because bradykinin has three Pro residues, two Arg residues, and two Phe residues.

In the amino acid analyzer, the components of the hydrolysate are dissolved in an aqueous buffer solution and separated by passing them down an ion-exchange column. The solution emerging from the column is mixed with ninhydrin, which reacts with amino acids to give the purple ninhydrin color. The absorption of light is recorded and printed out as a function of time.

The time required for each amino acid to pass through the column (its *retention time*) depends on how strongly that amino acid interacts with the ion-exchange resin. The retention time of each amino acid is known from standardization with pure amino acids. The amino acids present in the sample are identified by comparing their retention times with the known values. The area under each peak is nearly proportional to the amount of the amino acid producing that peak, so we can determine the relative amounts of amino acids present.

Figure 24-13 shows a standard trace of an equimolar mixture of amino acids, followed by the trace produced by the hydrolysate from human bradykinin (Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg).

**Sequencing the Peptide: Terminal Residue Analysis** The amino acid analyzer determines the amino acids present in a peptide, but it does not reveal their **sequence**: the order in which they are linked together. The peptide sequence is destroyed in the hydrolysis step. To determine the amino acid sequence, we must cleave just one amino acid from the chain and leave the rest of the chain intact. The cleaved amino acid can be separated and identified, and the process can be repeated on the rest of the chain. The amino acid may be cleaved from either end of the peptide (either the N terminus or the C terminus), and we will consider one method used for each end. This general method for peptide sequencing is called **terminal residue analysis**.

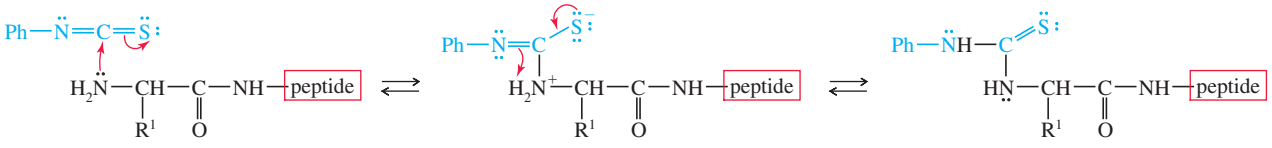
## 24-9C Sequencing from the N Terminus: The Edman Degradation

The most efficient method for sequencing peptides is the **Edman degradation**. A peptide is treated with phenyl isothiocyanate, followed by acid hydrolysis. The products are the shortened peptide chain and a heterocyclic derivative of the N-terminal amino acid called a *phenylthiohydantoin*.

This reaction takes place in three stages. First, the free amino group of the N-terminal amino acid reacts with phenylisothiocyanate to form a phenylthiourea.

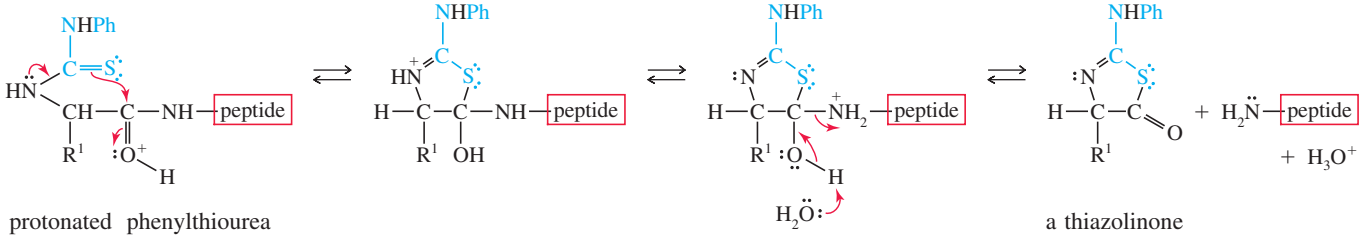
Second, the phenylthiourea cyclizes to a thiazolinone and expels the shortened peptide chain. Third, the thiazolinone isomerizes to the more stable phenylthiohydantoin.

*Step 1: Nucleophilic attack by the free amino group on phenyl isothiocyanate, followed by a proton transfer, gives a phenylthiourea.*



a phenylthiourea

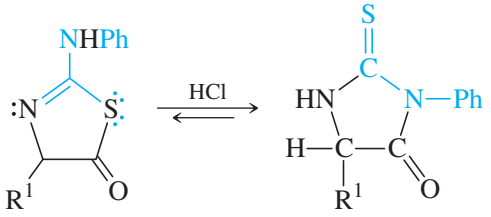
*Step 2: Treatment with HCl induces cyclization to a thiazolinone and expulsion of the shortened peptide chain.*



protonated phenylthiourea

a thiazolinone

*Step 3: In acid, the thiazolinone isomerizes to the more stable phenylthiohydantoin.*



thiazolinone

a phenylthiohydantoin

The phenylthiohydantoin derivative is identified by chromatography, by comparing it with phenylthiohydantoin derivatives of the standard amino acids. This gives the identity of the original N-terminal amino acid. The rest of the peptide is cleaved intact, and further Edman degradations are used to identify additional amino acids in the chain. This process is well suited to automation, and several types of automatic sequencers have been developed.

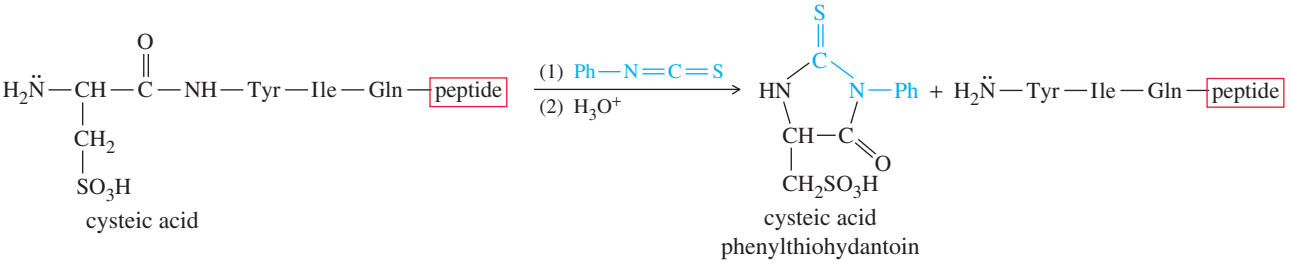
Figure 24-14 shows the first two steps in the sequencing of oxytocin. Before sequencing, the oxytocin sample is treated with peroxyformic acid to convert the disulfide bridge to cysteic acid residues.

In theory, Edman degradations could sequence a peptide of any length. In practice, however, the repeated cycles of degradation cause some internal hydrolysis of the peptide, with loss of sample and accumulation of by-products. After about 30 cycles of degradation, further accurate analysis becomes impossible. A small peptide such as bradykinin can be completely determined by Edman degradation, but larger proteins must be broken into smaller fragments (Section 24-9E) before they can be completely sequenced.

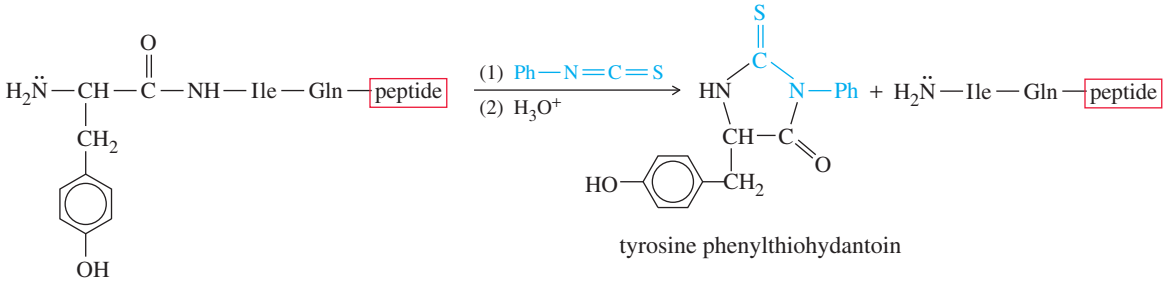
**PROBLEM 24-21**

Draw the structure of the phenylthiohydantoin derivatives of
**(a)** alanine **(b)** tryptophan **(c)** lysine **(d)** proline

*Step 1:* Cleavage and determination of the N-terminal amino acid




*Step 2:* Cleavage and determination of the second amino acid (the new N-terminal amino acid)




■ **FIGURE 24-14**
The first two steps in sequencing oxytocin. Each Edman degradation cleaves the N-terminal amino acid and forms its phenylthiohydantoin derivative. The shortened peptide is available for the next step.

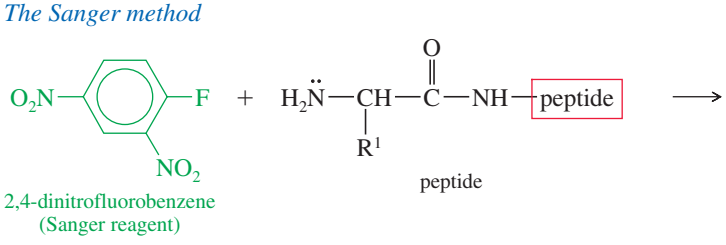
## PROBLEM 24-22

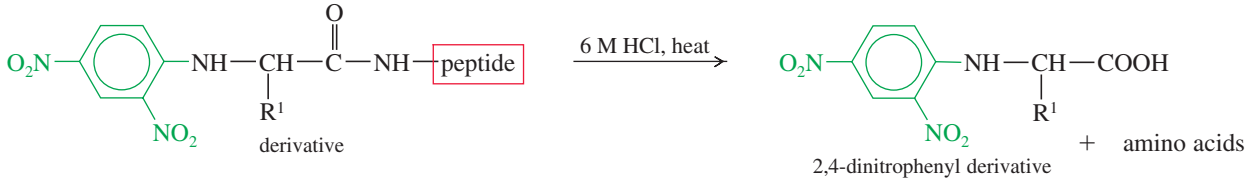
Show the third and fourth steps in the sequencing of oxytocin. Use Figure 24-14 as a guide.

## PROBLEM 24-23

The **Sanger method** for N-terminus determination is a less common alternative to the Edman degradation. In the Sanger method, the peptide is treated with the Sanger reagent, 2,4-dinitrofluorobenzene, and then hydrolyzed by reaction with 6 M aqueous HCl. The N-terminal amino acid is recovered as its 2,4-dinitrophenyl derivative and identified.

*The Sanger method*



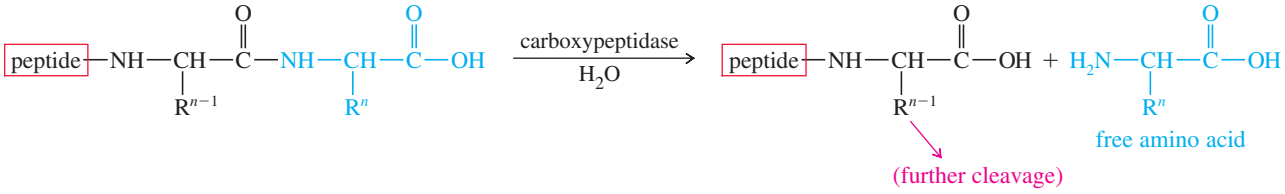




**(a)** Propose a mechanism for the reaction of the N terminus of the peptide with 2,4-dinitrofluorobenzene.
**(b)** Explain why the Edman degradation is usually preferred over the Sanger method.

## 24-9D C-Terminal Residue Analysis

There is no efficient method for sequencing several amino acids of a peptide starting from the C terminus. In many cases, however, the C-terminal amino acid can be identified using the enzyme *carboxypeptidase*, which cleaves the C-terminal peptide bond. The products are the free C-terminal amino acid and a shortened peptide. Further reaction cleaves the second amino acid that has now become the new C terminus of the shortened peptide. Eventually, the entire peptide is hydrolyzed to its individual amino acids.

The selective enzymatic cleavage of proteins is critical to many biological processes. For example, the clotting of blood depends on the enzyme thrombin cleaving fibrinogen at specific points to produce fibrin, the protein that forms a clot.




A peptide is incubated with the carboxypeptidase enzyme, and the appearance of free amino acids is monitored. In theory, the amino acid whose concentration increases first should be the C terminus, and the next amino acid to appear should be the second residue from the end. In practice, different amino acids are cleaved at different rates, making it difficult to determine amino acids past the C terminus and occasionally the second residue in the chain.

## 24-9E Breaking the Peptide into Shorter Chains: Partial Hydrolysis

Before a large protein can be sequenced, it must be broken into smaller chains, not longer than about 30 amino acids. Each of these shortened chains is sequenced, and then the entire structure of the protein is deduced by fitting the short chains together like pieces of a jigsaw puzzle.

Partial cleavage can be accomplished either by using dilute acid with a shortened reaction time or by using enzymes, such as *trypsin* and *chymotrypsin*, that break bonds between specific amino acids. The acid-catalyzed cleavage is not very selective, leading to a mixture of short fragments resulting from cleavage at various positions. Enzymes are more selective, giving cleavage at predictable points in the chain.

> TRYPSIN: Cleaves the chain at the carboxyl groups of the basic amino acids lysine and arginine.
> CHYMOTRYPSIN: Cleaves the chain at the carboxyl groups of the aromatic amino acids phenylalanine, tyrosine, and tryptophan.

Let's use oxytocin (Figure 24-9) as an example to illustrate the use of partial hydrolysis. Oxytocin could be sequenced directly by C-terminal analysis and a series of Edman degradations, but it provides a simple example of how a structure can be pieced together from fragments. Acid-catalyzed partial hydrolysis of oxytocin (after cleavage of the disulfide bridge) gives a mixture that includes the following peptides:

Ile-Gln-Asn-Cys  Gln-Asn-Cys-Pro  Pro-Leu-Gly · $NH_2$  Cys-Tyr-Ile-Gln-Asn  Cys-Pro-Leu-Gly

When we match the overlapping regions of these fragments, the complete sequence of oxytocin appears:

Cys-Tyr-Ile-Gln-Asn
Ile-Gln-Asn-Cys
Gln-Asn-Cys-Pro
Cys-Pro-Leu-Gly
Pro-Leu-Gly · $NH_2$

*Complete structure*

Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly · $NH_2$

Proteolytic (protein-cleaving) enzymes also have applications in consumer products. For example, papain (from papaya extract) serves as a meat tenderizer. It cleaves the fibrous proteins, making the meat less tough.

The two Cys residues in oxytocin may be involved in disulfide bridges, either linking two of these peptide units or forming a ring. By measuring the molecular weight of oxytocin, we can show that it contains just one of these peptide units; therefore, the Cys residues must link the molecule in a ring.

**PROBLEM 24-24**

Show where trypsin and chymotrypsin would cleave the following peptide.

Tyr-Ile-Gln-Arg-Leu-Gly-Phe-Lys-Asn-Trp-Phe-Gly-Ala-Lys-Gly-Gln-Gln · $NH_2$

**PROBLEM 24-25**

After treatment with peroxyformic acid, the peptide hormone vasopressin is partially hydrolyzed. The following fragments are recovered. Propose a structure for vasopressin.

| | | |
|---|---|---|
| Phe-Gln-Asn | Pro-Arg-Gly · $NH_2$ | Cys-Tyr-Phe |
| Asn-Cys-Pro-Arg | Tyr-Phe-Gln-Asn | |

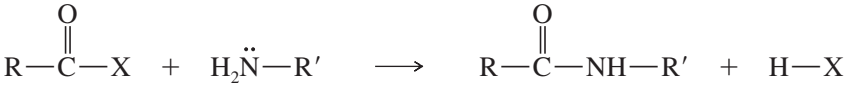
## 24-10 Solution-Phase Peptide Synthesis

### 24-10A Introduction

Total synthesis of peptides is rarely an economical method for their commercial production. Important peptides are usually derived from biological sources. For example, insulin for diabetics was originally taken from pork pancreas. Now, recombinant DNA techniques have improved the quality and availability of peptide pharmaceuticals. It is possible to extract the piece of DNA that contains the code for a particular protein, insert it into a bacterium, and induce the bacterium to produce the protein. Strains of *Escherichia coli* have been developed to produce human insulin that avoids dangerous reactions in people who are allergic to pork products.

Laboratory peptide synthesis is still an important area of chemistry, however, for two reasons: If the synthetic peptide is the same as the natural peptide, it proves the structure is correct; and the synthesis provides a larger amount of the material for further biological testing. Also, synthetic peptides can be made with altered amino acid sequences to compare their biological activity with the natural peptides. These comparisons can point out the critical areas of the peptide, which may suggest causes and treatments for genetic diseases involving similar abnormal peptides.

Peptide synthesis requires the formation of amide bonds between the proper amino acids in the proper sequence. With simple acids and amines, we would form an amide bond simply by converting the acid to an activated derivative (such as an acyl halide or anhydride) and adding the amine.



(X is a good leaving group, preferably electron-withdrawing)

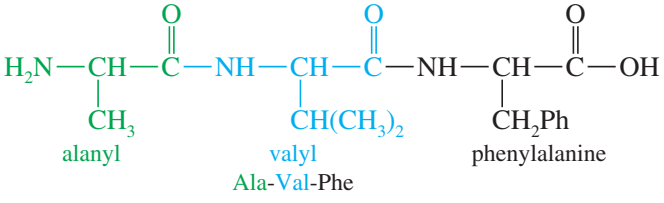
Amide formation is not so easy with amino acids, however. Each amino acid has both an amino group and a carboxyl group. If we activate the carboxyl group, it reacts with its own amino group. If we mix some amino acids and add a reagent to make them couple, they form every conceivable sequence. Also, some amino acids have side chains that might interfere with peptide formation. For example, glutamic acid has an extra carboxyl group, and lysine has an extra amino group. As a result, peptide synthesis always involves both activating reagents to form the correct peptide bonds and protecting groups to block formation of incorrect bonds.

Chemists have developed many ways of synthesizing peptides, falling into two major groups. The *solution-phase method* involves adding reagents to solutions

of growing peptide chains and purifying the products as needed. The *solid-phase method* involves adding reagents to growing peptide chains bonded to solid polymer particles. Many different reagents are available for each of these methods, but we will consider only one set of reagents for the solution-phase method and one set for the solid-phase method. The general principles are the same regardless of the specific reagents.

## 24-10B Solution-Phase Method

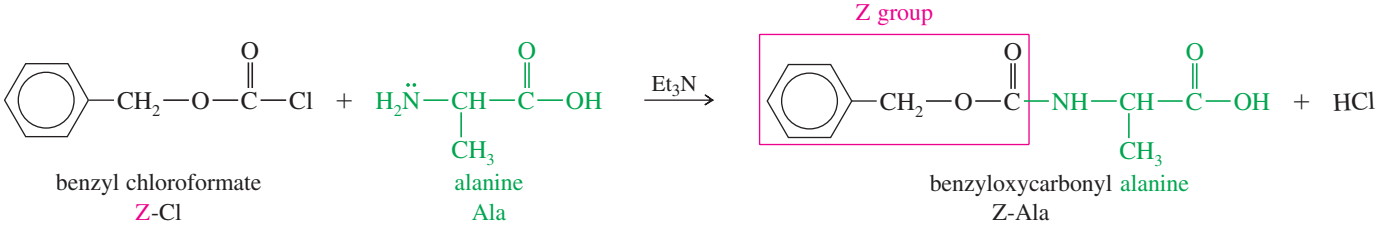
Consider the structure of alanylvalylphenylalanine, a simple tripeptide:




**Solution-phase peptide synthesis** begins at the N terminus and ends at the C terminus, or left to right as we draw the peptide. The first major step is to couple the carboxyl group of alanine to the amino group of valine. This cannot be done simply by activating the carboxyl group of alanine and adding valine. If we activated the carboxyl group of alanine, it would react with another molecule of alanine.

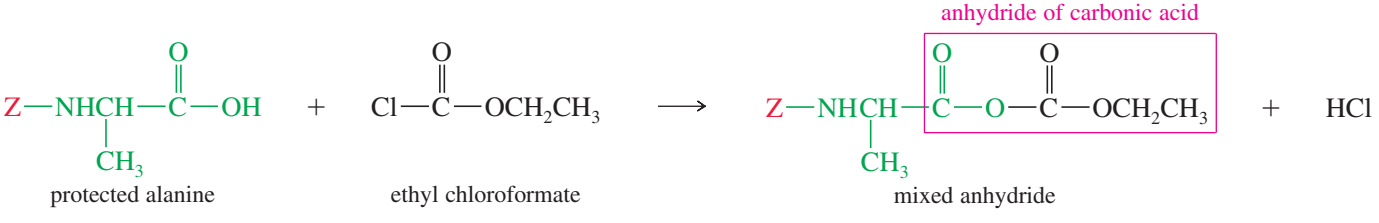
To prevent side reactions, the amino group of alanine must be protected to make it nonnucleophilic. In Section 24-7B, we saw that an amino acid reacts with benzyl chloroformate (also called *benzyloxycarbonyl chloride*) to form a urethane, or carbamate ester, that is easily removed at the end of the synthesis. This protecting group has been used for many years, and it has acquired several names. It is called the *benzyloxycarbonyl group*, the *carbobenzoxy group* (Cbz), or simply the *Z group* (abbreviated Z).

***Preliminary step:*** Protect the amino group with Z.



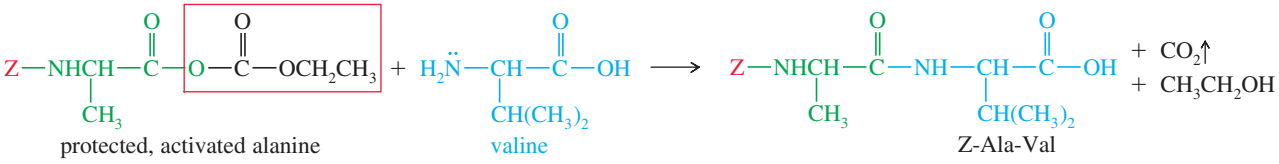

The amino group in Z-Ala is protected as the nonnucleophilic amide half of a carbamate ester. The carboxyl group can be activated without reacting with the protected amino group. Treatment with ethyl chloroformate converts the carboxyl group to a mixed anhydride of the amino acid and carbonic acid. It is strongly activated toward nucleophilic attack.

***Step 1:*** Activate the carboxyl group with ethyl chloroformate.



When the second amino acid (valine) is added to the protected, activated alanine, the nucleophilic amino group of valine attacks the activated carbonyl of alanine, displacing the anhydride and forming a peptide bond. (Some procedures use an ester of the new amino acid to avoid competing reactions from its carboxylate group.)

***Step 2:*** Form an amide bond to couple the next amino acid.

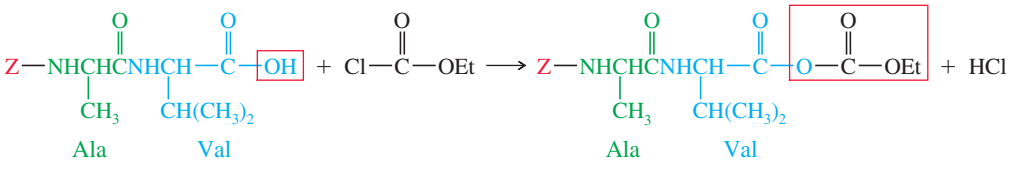



**PROBLEM 24-26**

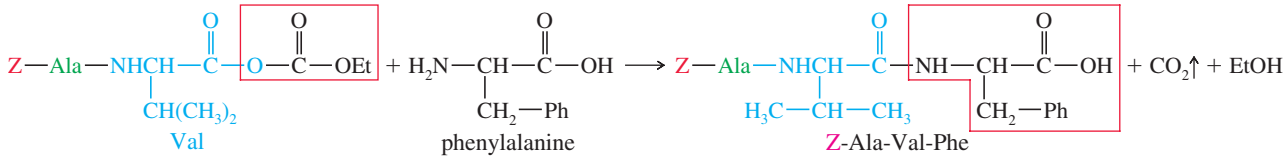
Give complete mechanisms for the formation of Z-Ala, its activation by ethyl chloroformate, and the coupling with valine.

At this point, we have the N-protected dipeptide Z-Ala-Val. Phenylalanine must be added to the C terminus to complete the Ala-Val-Phe tripeptide. Activation of the valine carboxyl group, followed by addition of phenylalanine, gives the protected tripeptide.

***Step 1:*** Activate the carboxyl group with ethyl chloroformate.




***Step 2:*** Form an amide bond to couple the next amino acid.

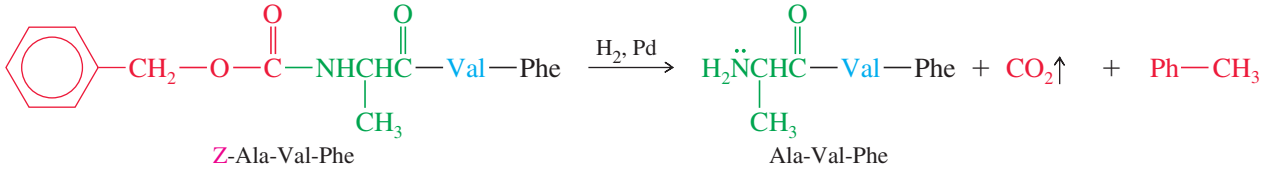



To make a larger peptide, repeat these two steps for the addition of each amino acid residue:

1. Activate the C terminus of the growing peptide by reaction with ethyl chloroformate.
2. Couple the next amino acid.

The final step in the solution-phase synthesis is to deprotect the N terminus of the completed peptide. The N-terminal amide bond must be cleaved without breaking any of the peptide bonds in the product. Fortunately, the benzyloxycarbonyl group is partly an amide and partly a benzyl ester, and hydrogenolysis of the benzyl ester takes place under mild conditions that do not cleave the peptide bonds. This mild cleavage is the reason for using the benzyloxycarbonyl group (as opposed to some other acyl group) to protect the N terminus.

***Final step:*** Remove the protecting group.



**PROBLEM 24-27**

Show how you would synthesize Ala-Val-Phe-Gly-Leu starting with Z-Ala-Val-Phe.

**PROBLEM 24-28**

Show how the solution-phase synthesis would be used to synthesize Ile-Gly-Asn.

***problem-solving*** **Hint**

Remember that classical (solution-phase) peptide synthesis:

1. Goes N → C. Protect the N terminus (Z group) first, deprotect last.
2. Couple each amino acid by activating the C terminus (ethyl chloroformate), then adding the new amino acid.

The solution-phase method works well for small peptides, and many peptides have been synthesized by this process. A large number of chemical reactions and purifications are required even for a small peptide, however. Although the individual yields are excellent, with a large peptide, the overall yield becomes so small as to be unusable, and several months (or years) are required to complete so many steps. The large amounts of time required and the low overall yields are due largely to the purification steps. For larger peptides and proteins, solid-phase peptide synthesis is usually preferred.

## 24-11 Solid-Phase Peptide Synthesis

In 1962, Robert Bruce Merrifield of Rockefeller University developed a method for synthesizing peptides without having to purify the intermediates. He did this by attaching the growing peptide chains to solid polystyrene beads. After each amino acid is added, the excess reagents are washed away by rinsing the beads with solvent. This ingenious method lends itself to automation, and Merrifield built a machine that can add several amino acid units while running unattended. Using this machine, Merrifield synthesized ribonuclease (124 amino acids) in just six weeks, obtaining an overall yield of 17%. Merrifield's work in solid-phase peptide synthesis won the Nobel Prize in Chemistry in 1984.

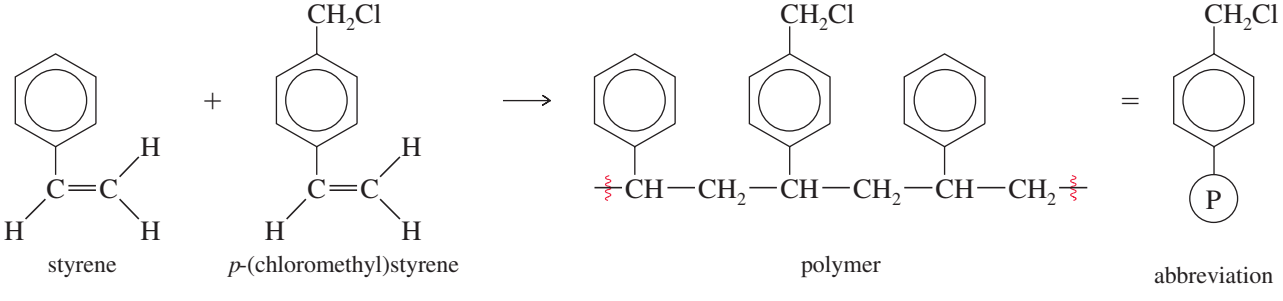
### 24-11A The Individual Reactions

Three reactions are crucial for solid-phase peptide synthesis. These reactions attach the first amino acid to the solid support, protect each amino group until its time to react, and form the peptide bonds between the amino acids.

**Attaching the Peptide to the Solid Support** The greatest difference between solution-phase and solid-phase peptide synthesis is that solid-phase synthesis is done in the opposite direction: starting with the C terminus and going toward the N terminus, right to left as we write the peptide. The first step is to attach the *last* amino acid (the C terminus) to the solid support.

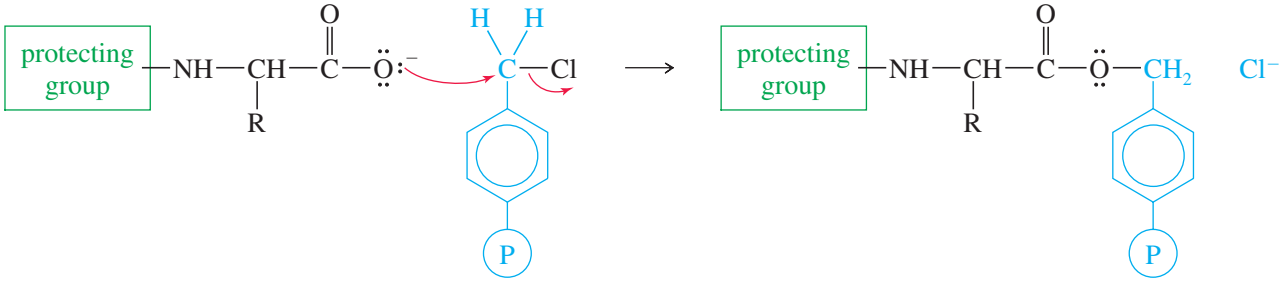
The solid support is a special polystyrene bead in which some of the aromatic rings have chloromethyl groups. This polymer, often called the *Merrifield resin*, is made by copolymerizing styrene with a few percent of *p*-(chloromethyl)styrene.

*Formation of the Merrifield resin*



Like other benzyl halides, the chloromethyl groups on the polymer are reactive toward $S_N2$ attack. The carboxyl group of an N-protected amino acid displaces chloride, giving an amino acid ester of the polymer. In effect, the polymer serves as the alcohol part of an ester protecting group for the carboxyl end of the C-terminal amino acid. The amino group must be protected, or it would attack the chloromethyl groups.

*Attachment of the C-terminal amino acid*

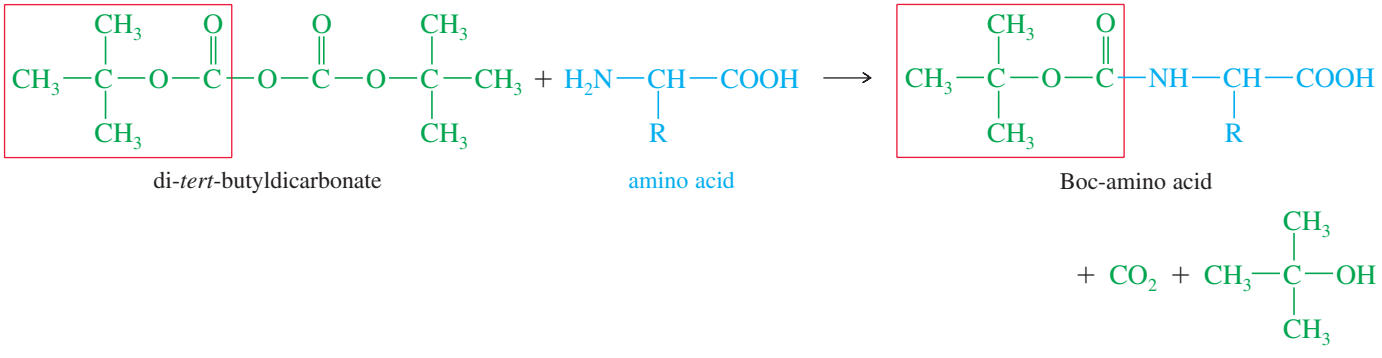



Once the C-terminal amino acid is fixed to the polymer, the chain is built on the amino group of this amino acid.

**Using the *tert*-Butyloxycarbonyl (Boc) Protecting Group** The benzyloxycarbonyl group (the Z group) cannot be used with the solid-phase process because the Z group is removed by hydrogenolysis in contact with a solid catalyst. A polymer-bound peptide cannot achieve the intimate contact with a solid catalyst required for hydrogenolysis. The N-protecting group used in the Merrifield procedure is the *tert*-butyloxycarbonyl group, abbreviated Boc or *t*-Boc. The Boc group is similar to the Z group, except that it has a *tert*-butyl group in place of the benzyl group. Like other *tert*-butyl esters, the Boc protecting group is easily removed under acidic conditions.

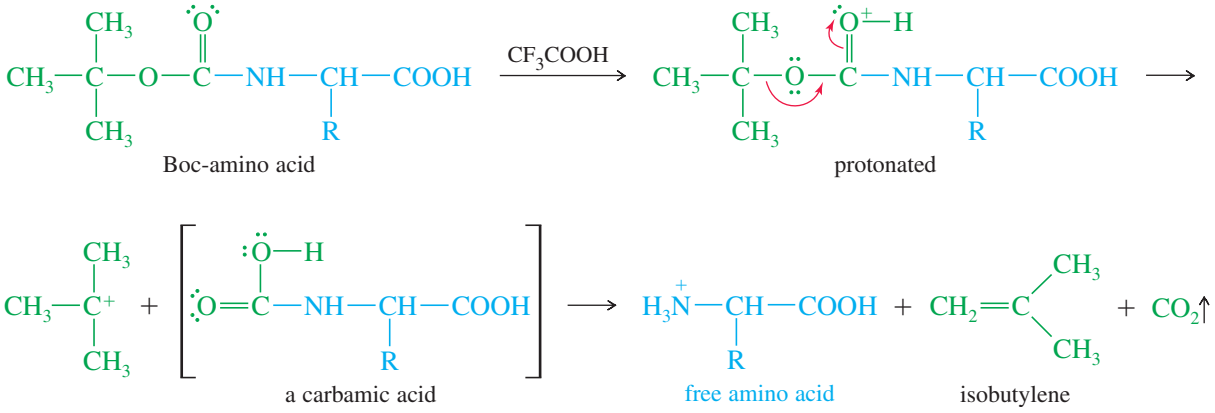
The acid chloride of the Boc group is unstable, so we use the anhydride, di-*tert*-butyldicarbonate, to attach the group to the amino acid.

*Protection of the amino group as its Boc derivative*



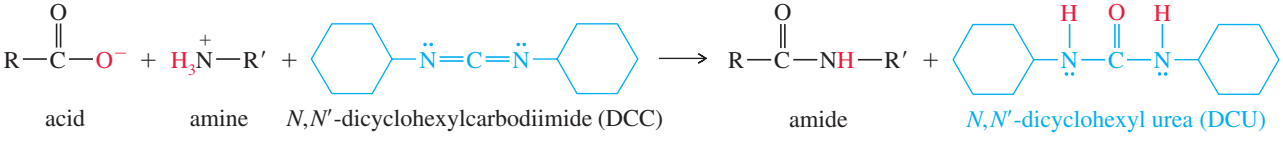

di-*tert*-butyldicarbonate amino acid Boc-amino acid

The Boc group is easily cleaved by brief treatment with trifluoroacetic acid (TFA), $CF_3COOH$. Loss of a relatively stable *tert*-butyl cation from the protonated ester gives an unstable carbamic acid. Decarboxylation of the carbamic acid gives the deprotected amino group of the amino acid. Loss of a proton from the *tert*-butyl cation gives isobutylene.



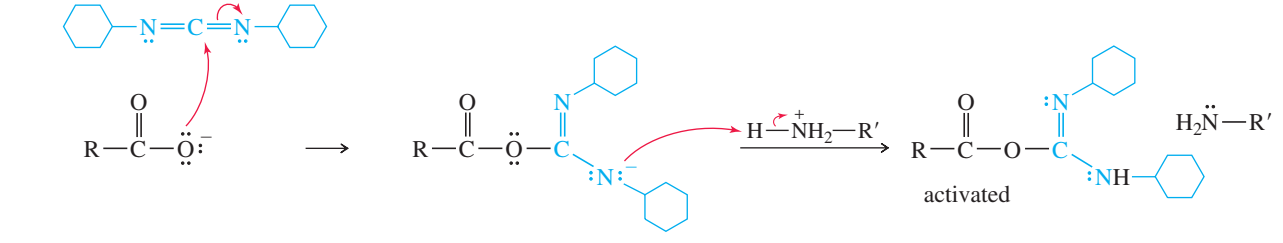

People who synthesize peptides generally do not make their own Boc-protected amino acids. Because they use all their amino acids in protected form, they buy and use commercially available Boc amino acids.

**Use of DCC as a Peptide Coupling Agent** The final reaction needed for the Merrifield procedure is the peptide bond-forming condensation. When a mixture of an amine and an acid is treated with *N*,*N*′-dicyclohexylcarbodiimide (abbreviated DCC), the amine and the acid couple to form an amide. The molecule of water lost in this condensation converts DCC to *N*,*N*′-dicyclohexyl urea (DCU).

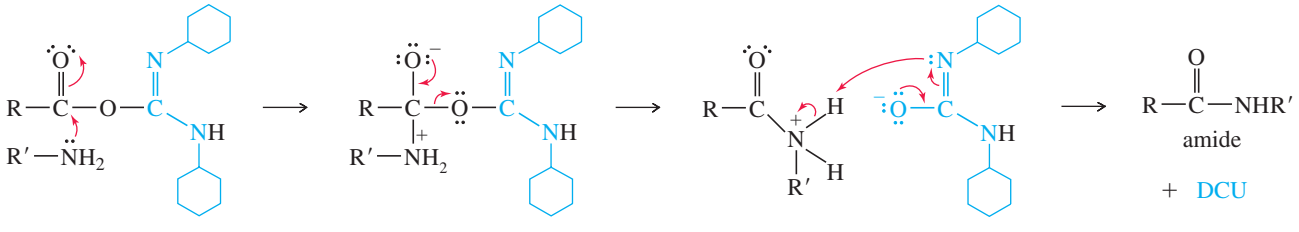



The mechanism for DCC coupling is not as complicated as it may seem. The carboxylate ion adds to the strongly electrophilic carbon of the diimide, giving an activated acyl derivative of the acid. This activated derivative reacts readily with the amine to give the amide. In the final step, DCU serves as an excellent leaving group. The cyclohexane rings are miniaturized for clarity.

*Formation of an activated acyl derivative*

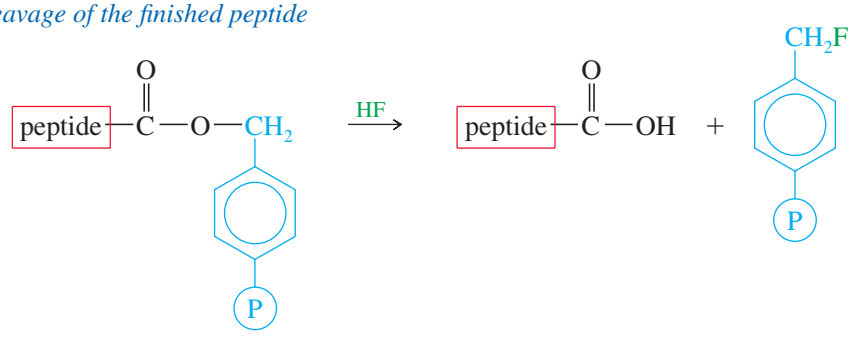



*Coupling with the amine and loss of DCU*



At the completion of the synthesis, the ester bond to the polymer is cleaved by anhydrous HF. Because this is an ester bond, it is more easily cleaved than the amide bonds of the peptide.

*Cleavage of the finished peptide*




**PROBLEM 24-29**

Propose a mechanism for the coupling of acetic acid and aniline using DCC as a coupling agent.

***problem-solving*** **Hint**

Remember that solid-phase peptide synthesis:

1. Goes C → N. Attach the Boc-protected C terminus to the bead first.
2. Couple each amino acid by removing (TFA) the Boc group from the N terminus, then add the next Boc-protected amino acid with DCC.
3. Cleave (HF) the finished peptide from the bead.

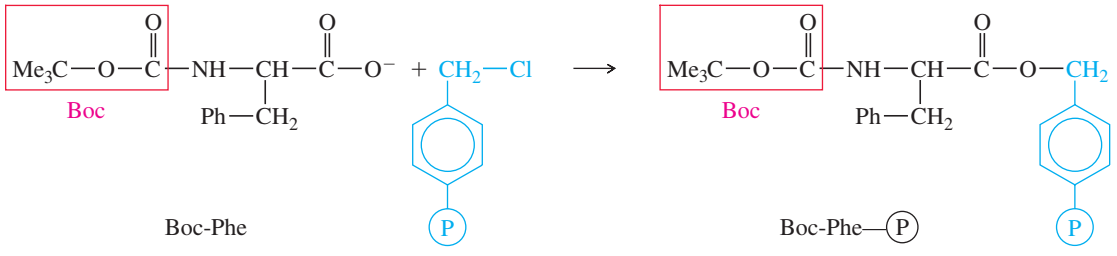
Now we consider an example to illustrate how these procedures are combined in the Merrifield solid-phase peptide synthesis.

## 24-11B An Example of Solid-Phase Peptide Synthesis

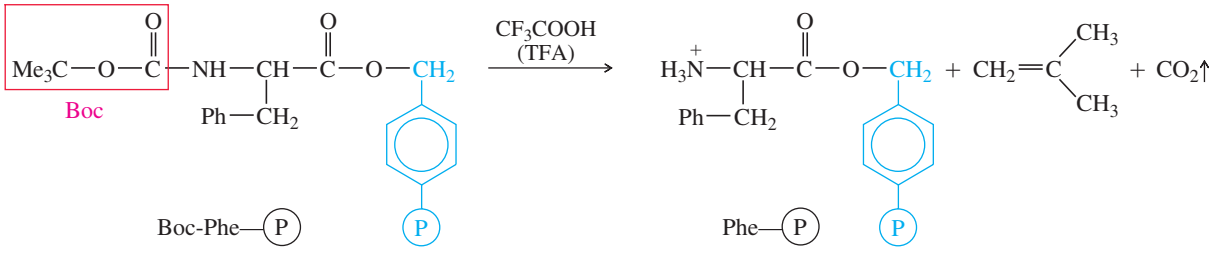
For easy comparison of the solution-phase and solid-phase methods, we will consider the synthesis of the same tripeptide we made using the solution-phase method.

Ala-Val-Phe

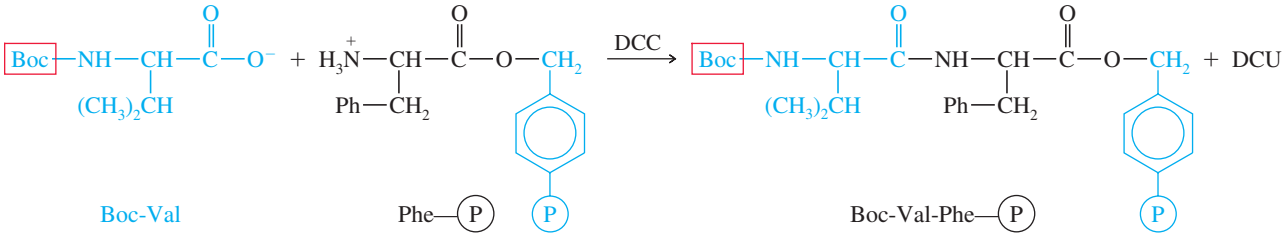
The solid-phase synthesis is carried out in the direction opposite that of the solution-phase synthesis. The first step is attachment of the N-protected C-terminal amino acid (Boc-phenylalanine) to the polymer.




Trifluoroacetic acid (TFA) cleaves the Boc protecting group of phenylalanine so that its amino group can be coupled with the next amino acid.

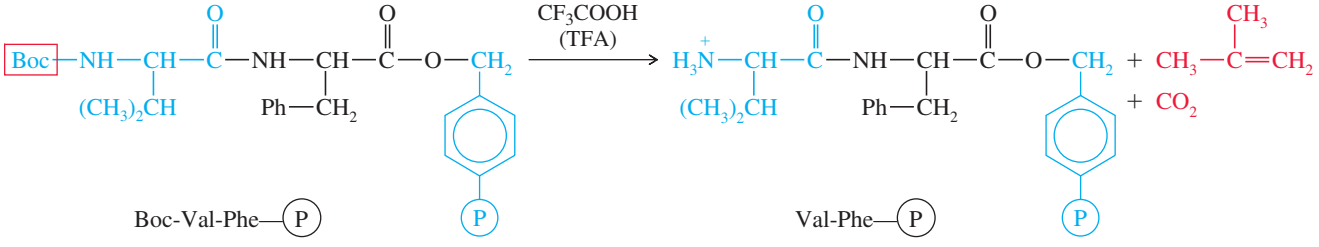


The second amino acid (valine) is added in its N-protected Boc form so that it cannot couple with itself. Addition of DCC couples the valine carboxyl group with the free $-NH_2$ group of phenylalanine.

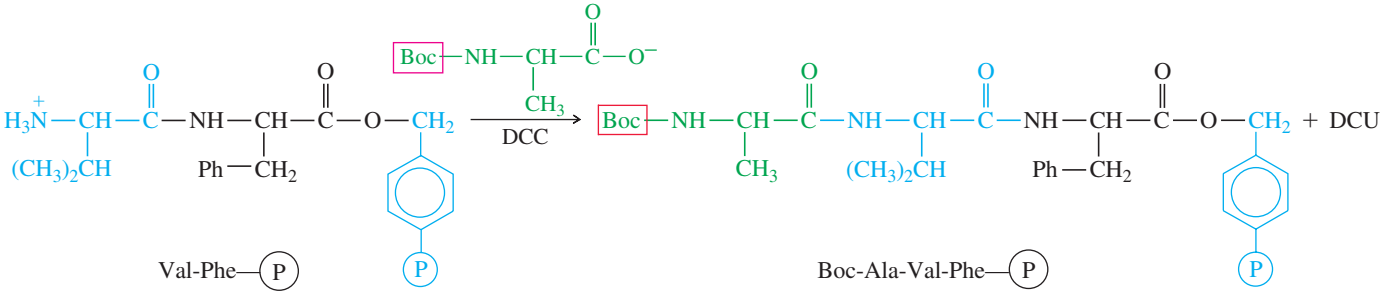



To couple the final amino acid (alanine), the chain is first deprotected by treatment with trifluoroacetic acid. Then the N-protected Boc-alanine and DCC are added.

*Step 1: Deprotection*



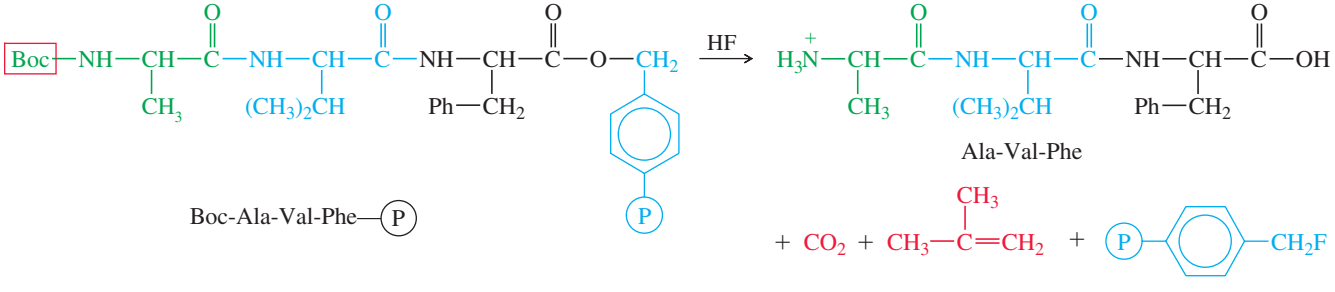

*Step 2: Coupling*




If we were making a longer peptide, the addition of each subsequent amino acid would require the repetition of two steps:

1. Use trifluoroacetic acid to deprotect the amino group at the end of the growing chain.
2. Add the next Boc-amino acid, using DCC as a coupling agent.

Once the peptide is completed, the final Boc protecting group must be removed, and the peptide must be cleaved from the polymer. Anhydrous HF cleaves the ester linkage that bonds the peptide to the polymer, and it also removes the Boc protecting group. In our example, the following reaction occurs:



**PROBLEM 24-30**

Show how you would synthesize Leu-Gly-Ala-Val-Phe starting with Boc-Ala-Val-Phe—Ⓟ.

**PROBLEM 24-31**

Show how solid-phase peptide synthesis would be used to make Ile-Gly-Asn.

## 24-12 Classification of Proteins

Proteins may be classified according to their chemical composition, their shape, or their function. Protein composition and function are treated in detail in a biochemistry course. For now, we briefly survey the types of proteins and their general classifications.

Proteins are grouped into *simple* and *conjugated* proteins according to their chemical composition. **Simple proteins** are those that hydrolyze to give only amino acids. All the protein structures we have considered so far are simple proteins. Examples are insulin, ribonuclease, oxytocin, and bradykinin. **Conjugated proteins** are bonded to a nonprotein **prosthetic group** such as a sugar, a nucleic acid, a lipid, or some other group. Table 24-3 lists some examples of conjugated proteins.

**TABLE 24-3**

Classes of Conjugated Proteins

| Class | Prosthetic Group | Examples |
|---|---|---|
| glycoproteins | carbohydrates | γ-globulin, interferon |
| nucleoproteins | nucleic acids | ribosomes, viruses |
| lipoproteins | fats, cholesterol | high-density lipoprotein |
| metalloproteins | a complexed metal | hemoglobin, cytochromes |

Proteins are classified as *fibrous* or *globular* depending on whether they form long filaments or coil up on themselves. **Fibrous proteins** are stringy, tough, and usually insoluble in water. They function primarily as structural parts of the organism. Examples of fibrous proteins are α-keratin in hooves and fingernails, and collagen in tendons. **Globular proteins** are folded into roughly spherical shapes. They usually function as enzymes, hormones, or transport proteins. **Enzymes** are protein-containing biological catalysts; an example is ribonuclease, which cleaves RNA. Hormones help to regulate processes in the body. An example is insulin, which regulates glucose levels in the blood and its uptake by cells. Transport proteins bind to specific molecules and transport them in the blood or through the cell membrane. An example is hemoglobin, which transports oxygen in the blood from the lungs to the tissues.

## 24-13 Levels of Protein Structure

### 24-13A Primary Structure

Up to now, we have discussed the *primary structure* of proteins. The **primary structure** is the covalently bonded structure of the molecule. This definition includes the sequence of amino acids, together with any disulfide bridges. All the properties of the protein are determined, directly or indirectly, by the primary structure. Any folding, hydrogen bonding, or catalytic activity depends on the proper primary structure.

### 24-13B Secondary Structure

Although we often think of peptide chains as linear structures, they tend to form orderly hydrogen-bonded arrangements. In particular, the carbonyl oxygen atoms form hydrogen

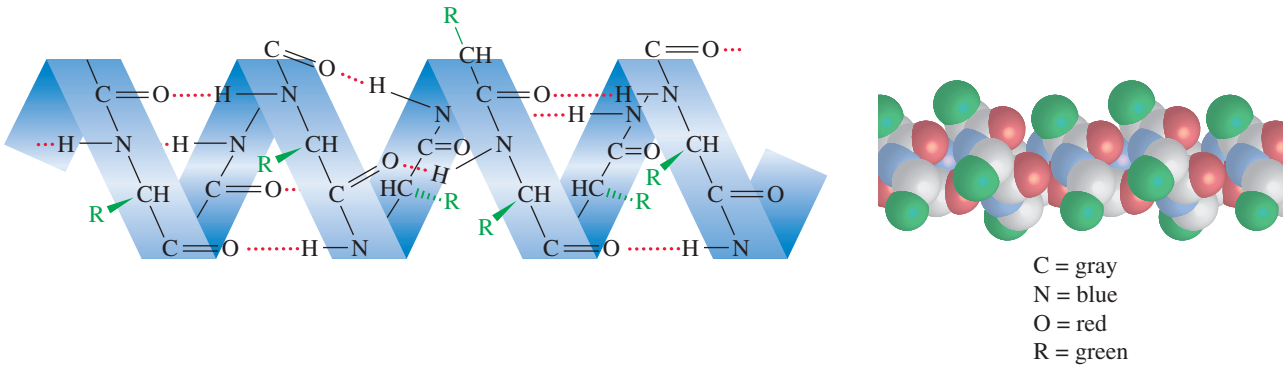



**FIGURE 24-15**
The $\alpha$ helical arrangement. The peptide chain curls into a helix so that each peptide carbonyl group is hydrogen-bonded to an N—H hydrogen on the next turn of the helix. Side chains are symbolized by green atoms in the space-filling structure.

bonds with the amide (N—H) hydrogens. This tendency leads to orderly patterns of hydrogen bonding: the ***α* helix** and the **pleated sheet**. These hydrogen-bonded arrangements, if present, are called the **secondary structure** of the protein.

When a peptide chain winds into a helical coil, each carbonyl oxygen can hydrogen-bond with an N—H hydrogen on the next turn of the coil. Many proteins wind into an $\alpha$ helix (a helix that looks like the thread on a right-handed screw) with the side chains positioned on the outside of the helix. For example, the fibrous protein $\alpha$ keratin is arranged in the $\alpha$-helical structure, and most globular proteins contain segments of $\alpha$ helix. Figure 24-15 shows the $\alpha$-helical arrangement.

Segments of peptides can also form orderly arrangements of hydrogen bonds by lining up side-by-side. In this arrangement, each carbonyl group on one chain forms a hydrogen bond with an N—H hydrogen on an adjacent chain. This arrangement may involve many peptide molecules lined up side-by-side, resulting in a two-dimensional *sheet*. The bond angles between amino acid units are such that the sheet is *pleated* (creased), with the amino acid side chains arranged on alternating sides of the sheet. Silk fibroin, the principal fibrous protein in the silks of insects and arachnids, has a pleated sheet secondary structure. Figure 24-16 shows the pleated sheet structure.

Spider web is composed mostly of fibroin, a protein with pleated-sheet secondary structure. The pleated-sheet arrangement allows for multiple hydrogen bonds between molecules, conferring great strength.

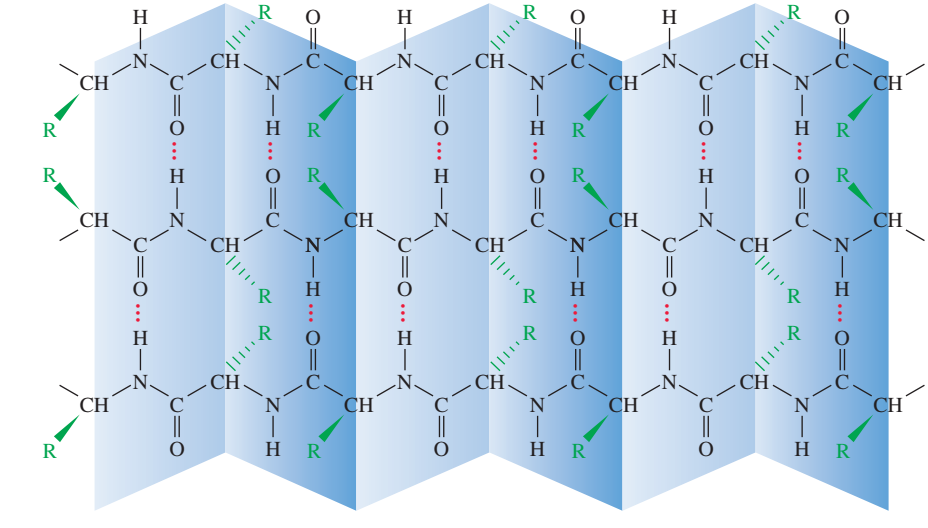


**FIGURE 24-16**
The pleated sheet arrangement. Each peptide carbonyl group is hydrogen-bonded to an N—H hydrogen on an adjacent peptide chain.

Tertiary structures of proteins are determined by X-ray crystallography. A single crystal of the protein is bombarded with X rays, whose wavelengths are appropriate to be diffracted by the regular atomic spacings in the crystal. A computer then determines the locations of the atoms in the crystal.

A protein may or may not have the same secondary structure throughout its length. Some parts may be curled into an $\alpha$ helix, while other parts are lined up in a pleated sheet. Parts of the chain may have no orderly secondary structure at all. Such a structureless region is called a **random coil**. Most globular proteins, for example, contain segments of $\alpha$ helix or pleated sheet separated by kinks of random coil, allowing the molecule to fold into its globular shape.

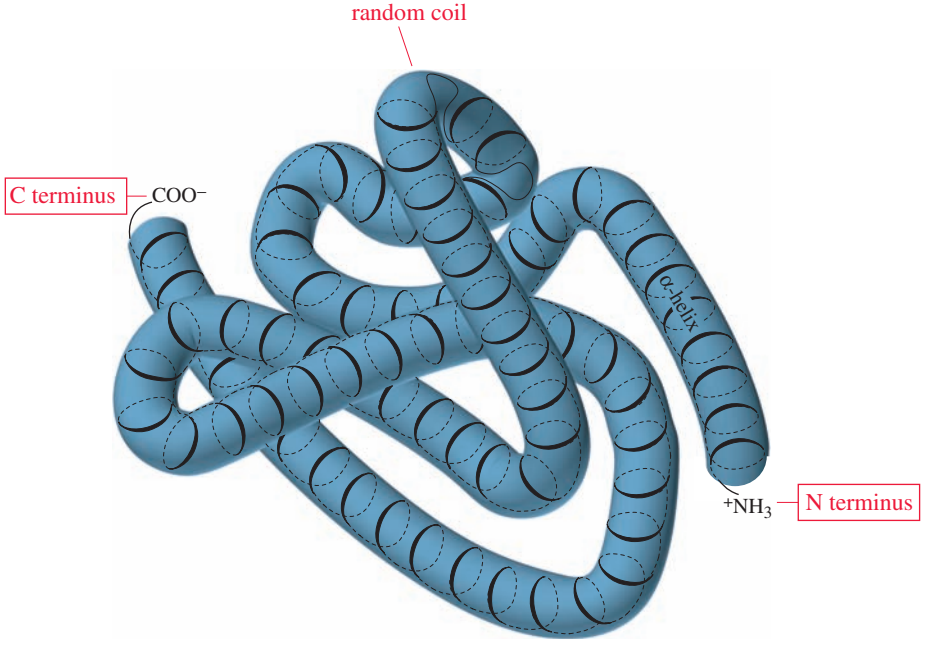
## 24-13C Tertiary Structure

The **tertiary structure** of a protein is its complete three-dimensional conformation. Think of the secondary structure as a spatial pattern in a local region of the molecule. Parts of the protein may have the $\alpha$-helical structure, while other parts may have the pleated-sheet structure, and still other parts may be random coils. The tertiary structure includes all the secondary structure and all the kinks and folds in between. The tertiary structure of a typical globular protein is represented in Figure 24-17.

Coiling of an enzyme can give three-dimensional shapes that produce important catalytic effects. Polar, *hydrophilic* (water-loving) side chains are oriented toward the outside of the globule. Nonpolar, *hydrophobic* (water-hating) groups are arranged toward the interior. Coiling in the proper conformation creates an enzyme's **active site**, the region that binds the substrate and catalyzes the reaction. A reaction taking place at the active site in the interior of an enzyme may occur under essentially anhydrous, nonpolar conditions—while the whole system is dissolved in water!

## 24-13D Quaternary Structure

**Quaternary structure** refers to the association of two or more peptide chains in the complete protein. Not all proteins have quaternary structure. The ones that do are those that associate together in their active form. For example, hemoglobin, the oxygen carrier in mammalian blood, consists of four peptide chains fitted together to form a globular protein. Figure 24-18 summarizes the four levels of protein structure.




■ **FIGURE 24-17**
The tertiary structure of a typical globular protein includes segments of $\alpha$ helix with segments of random coil at the points where the helix is folded.

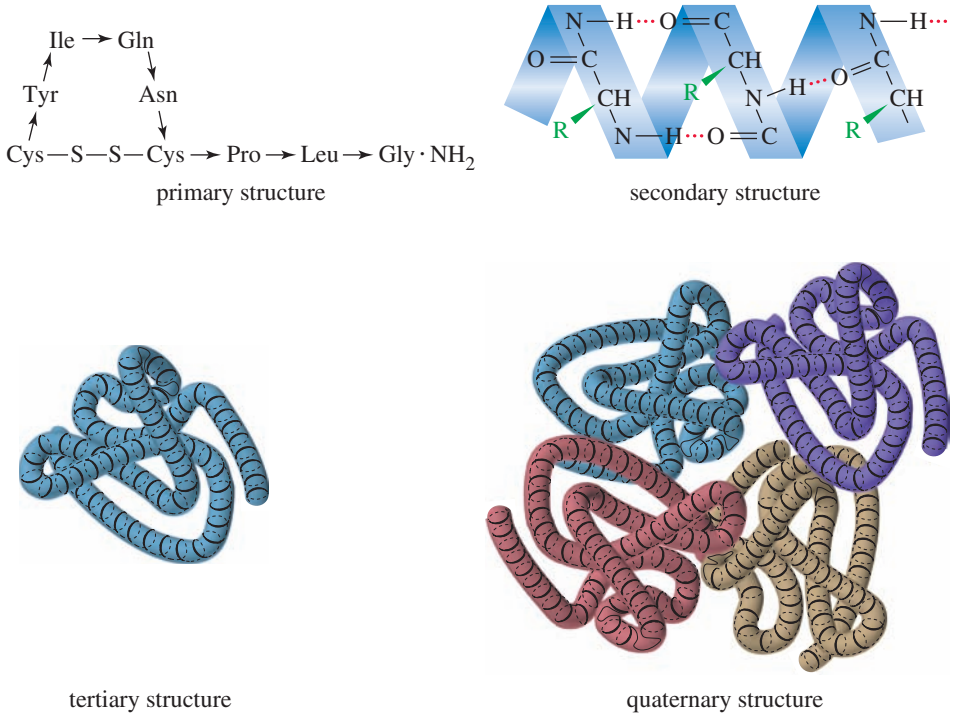



**FIGURE 24-18**
A schematic comparison of the levels of protein structure. Primary structure is the covalently bonded structure, including the amino acid sequence and any disulfide bridges. Secondary structure refers to the areas of $\alpha$ helix, pleated sheet, or random coil. Tertiary structure refers to the overall conformation of the molecule. Quaternary structure refers to the association of two or more peptide chains in the active protein.

## 24-14 Protein Denaturation

For a protein to be biologically active, it must have the correct structure at all levels. The sequence of amino acids must be right, with the correct disulfide bridges linking the cysteines on the chains. The secondary and tertiary structures are important, as well. The protein must be folded into its natural conformation, with the appropriate areas of $\alpha$ helix and pleated sheet. For an enzyme, the active site must have the right conformation, with the necessary side-chain functional groups in the correct positions. Conjugated proteins must have the right prosthetic groups, and multichain proteins must have the right combination of individual peptides.

With the exception of the covalent primary structure, all these levels of structure are maintained by weak solvation and hydrogen-bonding forces. Small changes in the environment can cause a chemical or conformational change resulting in **denaturation**: disruption of the normal structure and loss of biological activity. Many factors can cause denaturation, but the most common ones are heat and pH.

### 24-14A Reversible and Irreversible Denaturation

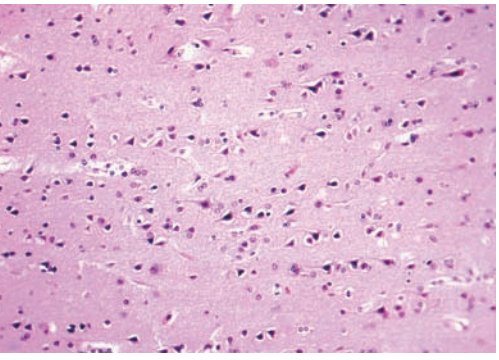
The cooking of egg white is an example of protein denaturation by high temperature. Egg white contains soluble globular proteins called *albumins*. When egg white is heated, the albumins unfold and coagulate to produce a solid rubbery mass. Different proteins have different abilities to resist the denaturing effect of heat. Egg albumin is quite sensitive to heat, but bacteria that live in geothermal hot springs have developed proteins that retain their activity in boiling water.

When a protein is subjected to an acidic pH, some of the side-chain carboxyl groups become protonated and lose their ionic charge. Conformational changes result, leading to denaturation. In a basic solution, amino groups become deprotonated, similarly losing their ionic charge, causing conformational changes and denaturation.

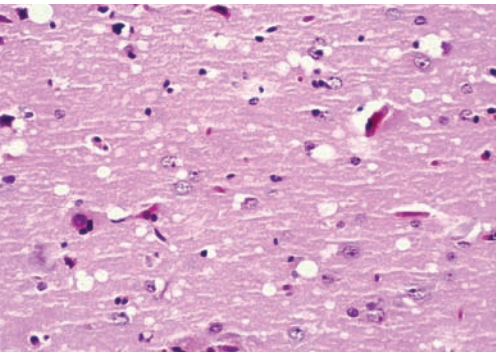
Irreversible denaturation of egg albumin. The egg white does not become clear and runny again when it cools.

Milk turns sour because of the bacterial conversion of carbohydrates to lactic acid. When the pH becomes strongly acidic, soluble proteins in milk are denatured and precipitate. This process is called *curdling*. Some proteins are more resistant to acidic and basic conditions than others. For example, most digestive enzymes such as amylase and trypsin remain active under acidic conditions in the stomach, even at a pH of about 1.

In many cases, denaturation is irreversible. When cooked egg white is cooled, it does not become uncooked. Curdled milk does not uncurdle when it is neutralized. Denaturation may be reversible, however, if the protein has undergone only mild denaturing conditions. For example, a protein can be *salted out* of solution by a high salt concentration, which denatures and precipitates the protein. When the precipitated protein is redissolved in a solution with a lower salt concentration, it usually regains its activity together with its natural conformation.



Micrograph of normal human brain tissue. The nuclei of neurons appear as dark spots.



Brain tissue of a patient infected with vCJD. Note the formation of (white) vacuole spaces and (dark, irregular) plaques of prion protein. (Magnification 200X)

## 24-14B Prion Diseases

Up through 1980, people thought that all infectious diseases were caused by microbes of some sort. They knew about diseases caused by viruses, bacteria, protozoa, and fungi. There were some strange diseases, however, for which no one had isolated and cultured the pathogen. *Creutzfeldt–Jakob Disease* (CJD) in humans, *scrapie* in sheep, and *transmissible encephalopathy* in mink (TME) all involved a slow, gradual loss of mental function and eventual death. The brains of the victims all showed unusual plaques of amyloid protein surrounded by spongelike tissue.

Workers studying these diseases thought there was an infectious agent involved (as opposed to genetic or environmental causes) because they knew that scrapie and TME could be spread by feeding healthy animals the ground-up remains of sick animals. They had also studied *kuru*, a disease much like CJD among tribes where family members showed their respect for the dead by eating their brains. These diseases were generally attributed to "slow viruses" that were yet to be isolated.

In the 1980s, neurologist Stanley B. Prusiner (of the University of California at San Francisco) made a homogenate of scrapie-infected sheep brains and systematically separated out all the cell fragments, bacteria, and viruses, and found that the remaining material was still infectious. He separated out the proteins and found a protein fraction that was still infectious. He suggested that scrapie (and presumably similar diseases) is caused by a protein infectious agent that he called **prion protein**. This conclusion contradicted the established principle that contagious diseases require a living pathogen. Many skeptical workers repeated Prusiner's work in hopes of finding viral contaminants in the infectious fractions, and most of them finally came to the same conclusion. Prusiner received the 1998 Nobel Prize in Medicine or Physiology for this work.

Since Prusiner's work, prion diseases have become more important because of their threat to humans. Beginning in 1996, some cows in the United Kingdom developed "mad cow disease" and would threaten other animals, wave their heads, fall down, and eventually die. The disease, called *bovine spongiform encephalopathy*, or BSE, was probably transmitted to cattle by feeding them the remains of scrapie-infected sheep. The most frightening aspect of the BSE outbreak was that people could contract a fatal disease, called *new-variant Creutzfeldt–Jakob Disease* (vCJD) from eating the infected meat. Since that time, a similar disease, called *chronic wasting disease*, or CWD, has been found in wild deer and elk in the Rocky Mountains. All of these (presumed) prion diseases are now classified as *transmissible spongiform encephalopathies*, or TSEs.

The most widely accepted theory of prion diseases suggests that the infectious prion protein has the same primary structure as a normal protein found in nerve

cells, but it differs in its tertiary structure. In effect, it is a misfolded, denatured version of a normal protein that polymerizes to form the amyloid protein plaques seen in the brains of infected animals. When an animal ingests infected food, the polymerized protein resists digestion. Because it is simply a misfolded version of a normal protein, the infectious prion does not provoke the host's immune system to attack the pathogen.

When the abnormal prion interacts with the normal version of the protein on the membranes of nerve cells, the abnormal protein somehow induces the normal molecules to change their shape. This is the part of the process we know the least about. (We might think of it like crystallization, in which a seed crystal induces other molecules to crystallize in the same conformation and crystal form.) These newly misfolded protein molecules then induce more molecules to change shape. The polymerized abnormal protein cannot be broken down by the usual protease enzymes, so it builds up in the brain and causes the plaques and spongy tissue associated with TSEs.

We once thought that a protein with the correct primary structure, placed in the right physiological solution, would naturally fold into the correct tertiary structure and stay that way. We were wrong. We now know that protein folding is a carefully controlled process in which enzymes and *chaperone proteins* promote correct folding as the protein is synthesized. Prion diseases have shown that there are many factors that cause proteins to fold into natural or unnatural conformations, and that the folding of the protein can have major effects on its biological properties within an organism.

# Glossary 24

**active site** The region of an enzyme that binds the substrate and catalyzes the reaction. (p. 1190)

**amino acid** Literally, any molecule containing both an amino group ($-NH_2$) and a carboxyl group ($-COOH$). The term usually means an **$\alpha$-amino acid**, with the amino group on the carbon atom next to the carboxyl group. (p. 1154)

**biomimetic synthesis** A laboratory synthesis that is patterned after a biological synthesis. For example, the synthesis of amino acids by reductive amination resembles the biosynthesis of glutamic acid. (p. 1162)

**complete proteins** Proteins that provide all the essential amino acids in about the right proportions for human nutrition. Examples include those in meat, fish, milk, and eggs. **Incomplete proteins** are severely deficient in one or more of the essential amino acids. Most plant proteins are incomplete. (p. 1157)

**conjugated protein** A protein that contains a nonprotein prosthetic group such as a sugar, nucleic acid, lipid, or metal ion. (p. 1188)

**C terminus (C-terminal end)** The end of the peptide chain with a free or derivatized carboxyl group. As the peptide is written, the C terminus is usually on the right. The amino group of the C-terminal amino acid links it to the rest of the peptide. (p. 1172)

**denaturation** An unnatural alteration of the conformation or the ionic state of a protein. Denaturation generally results in precipitation of the protein and loss of its biological activity. Denaturation may be reversible, as in salting out a protein, or irreversible, as in cooking an egg. (p. 1191)

**disulfide linkage (disulfide bridge)** A bond between two cysteine residues formed by mild oxidation of their thiol groups to a disulfide. (p. 1172)

**Edman degradation** A method for removing and identifying the N-terminal amino acid from a peptide without destroying the rest of the peptide chain. The peptide is treated with phenylisothiocyanate, followed by a mild acid hydrolysis to convert the N-terminal amino acid to its phenylthiohydantoin derivative. The Edman degradation can be used repeatedly to determine the sequence of many residues beginning at the N terminus. (p. 1176)

**electrophoresis** A procedure for separating charged molecules by their migration in a strong electric field. The direction and rate of migration are governed largely by the average charge on the molecules. (p. 1160)

**enzymatic resolution** The use of enzymes to separate enantiomers. For example, the enantiomers of an amino acid can be acylated and then treated with hog kidney acylase. The enzyme hydrolyzes the acyl group from the natural L-amino acid, but it does not react with the D-amino acid. The resulting mixture of the free L-amino acid and the acylated D-amino acid is easily separated. (p. 1167)

**enzyme** A protein-containing biological catalyst. Many enzymes also include *prosthetic groups*, nonprotein constituents that are essential to the enzyme's catalytic activity. (p. 1188)

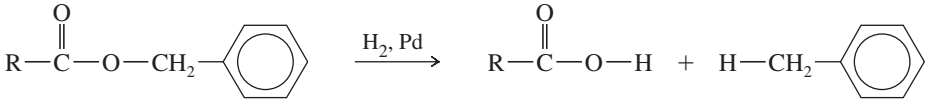
**essential amino acids** Ten standard amino acids that are not biosynthesized by humans and must be provided in the diet. (p. 1157)

**fibrous proteins** A class of proteins that are stringy, tough, threadlike, and usually insoluble in water. (p. 1188)

**globular proteins** A class of proteins that are relatively spherical in shape. Globular proteins generally have lower molecular weights and are more soluble in water than fibrous proteins. (p. 1188)

**α helix** A helical peptide conformation in which the carbonyl groups on one turn of the helix are hydrogen-bonded to N—H hydrogens on the next turn. Extensive hydrogen bonding stabilizes this helical arrangement. (p. 1189)

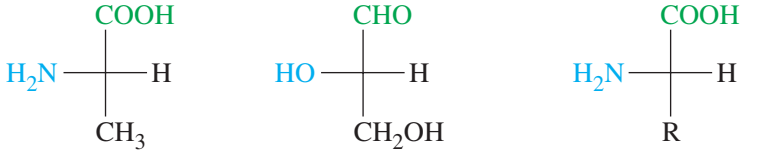
**hydrogenolysis** Cleavage of a bond by the addition of hydrogen. For example, catalytic hydrogenolysis cleaves benzyl esters. (p. 1168)




benzyl ester — acid — toluene

**isoelectric point (isoelectric pH)** The pH at which an amino acid (or protein) does not move under electrophoresis. This is the pH where the average charge on its molecules is zero, with most of the molecules in their zwitterionic form. (p. 1160)

**L-amino acid** An amino acid having a stereochemical configuration similar to that of L-(−)-glyceraldehyde. Most naturally occurring amino acids have the L configuration. (p. 1155)




L-alanine
(*S*)-alanine

L-(–)-glyceraldehyde
(*S*)-glyceraldehyde

an L-amino acid
(*S*) configuration

**N terminus (N-terminal end)** The end of the peptide chain with a free or derivatized amino group. As the peptide is written, the N terminus is usually on the left. The carboxyl group of the N-terminal amino acid links it to the rest of the peptide. (p. 1172)

**oligopeptide** A small polypeptide, containing about four to ten amino acid residues. (p. 1171)

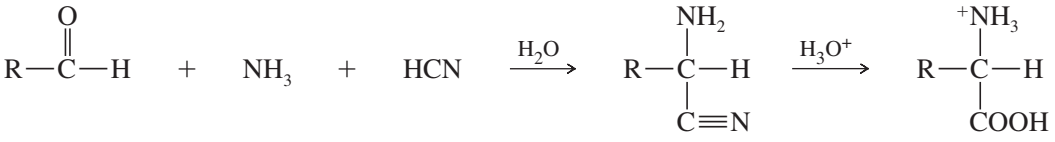
**peptide** Any polymer of amino acids linked by amide bonds between the amino group of each amino acid and the carboxyl group of the neighboring amino acid. The terms *dipeptide, tripeptide*, etc. may specify the number of amino acids in the peptide. (p. 1171)

**peptide bonds** Amide linkages between amino acids. (pp. 1153, 1171)

**pleated sheet** A two-dimensional peptide conformation with the peptide chains lined up side by side. The carbonyl groups on each peptide chain are hydrogen-bonded to N—H hydrogens on the adjacent chain, and the side chains are arranged on alternating sides of the sheet. (p. 1189)

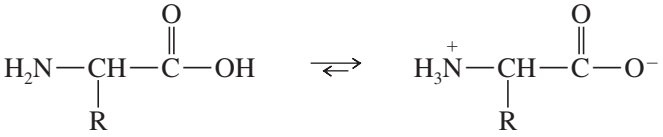
**polypeptide** A peptide containing many amino acid residues. Although proteins are polypeptides, the term *polypeptide* is commonly used for molecules with lower molecular weights than proteins. (p. 1171)

**primary structure** The covalently bonded structure of a protein; the sequence of amino acids, together with any disulfide bridges. (p. 1188)
**prion protein** A protein infectious agent that is thought to promote misfolding and polymerization of normal protein molecules, leading to amyloid plaques and destruction of nerve tissue. (p. 1192)
**prosthetic group** The nonprotein part of a conjugated protein. Examples of prosthetic groups are sugars, lipids, nucleic acids, and metal complexes. (p. 1188)
**protein** A biopolymer of amino acids. Proteins are polypeptides with molecular weights higher than about 6000 amu. (p. 1171)
**quaternary structure** The association of two or more peptide chains into a composite protein. (p. 1190)
**random coil** A type of protein secondary structure where the chain is neither curled into an $\alpha$-helix nor lined up in a pleated sheet. In a globular protein, the kinks that fold the molecule into its globular shape are usually segments of random coil. (p. 1190)
**residue** An amino acid unit of a peptide. (p. 1171)
**Sanger method** A method for determining the N-terminal amino acid of a peptide. The peptide is treated with 2,4-dinitrofluorobenzene (Sanger's reagent), then completely hydrolyzed. The derivatized amino acid is easily identified, but the rest of the peptide is destroyed in the hydrolysis. (p. 1178)
**secondary structure** The local hydrogen-bonded arrangement of a protein. The secondary structure is generally the $\alpha$ helix, pleated sheet, or random coil. (p. 1189)
**sequence** As a noun, the order in which amino acids are linked together in a peptide. As a verb, to determine the sequence of a peptide. (p. 1176)
**simple proteins** Proteins composed of only amino acids (having no prosthetic groups). (p. 1188)
**solid-phase peptide synthesis** A method in which the C-terminal amino acid is attached to a solid support (polystyrene beads) and the peptide is synthesized in the C → N direction by successive coupling of protected amino acids. When the peptide is complete, it is cleaved from the solid support. (p. 1183)
**solution-phase peptide synthesis (classical peptide synthesis)** Any of several methods in which protected amino acids are coupled in solution in the correct sequence to give a desired peptide. Most of these methods proceed in the N → C direction (p. 1181)
**standard amino acids** The 20 $\alpha$-amino acids found in nearly all naturally occurring proteins. (p. 1155)
**Strecker synthesis** Synthesis of $\alpha$-amino acids by reaction of an aldehyde with ammonia and cyanide ion, followed by hydrolysis of the intermediate $\alpha$-amino nitrile. (p. 1165)



aldehyde — $\alpha$-amino nitrile — $\alpha$-amino acid

**terminal residue analysis** Sequencing a peptide by removing and identifying the residue at the N terminus or at the C terminus. (p. 1176)
**tertiary structure** The complete three-dimensional conformation of a protein. (p. 1190)
**transamination** Transfer of an amino group from one molecule to another. Transamination is a common method for the biosynthesis of amino acids, often involving glutamic acid as the source of the amino group. (p. 1162)
**zwitterion (dipolar ion)** A structure with an overall charge of zero but having a positively charged substituent and a negatively charged substituent. Most amino acids exist in zwitterionic forms. (p. 1158)



uncharged structure (minor component) — dipolar ion, or zwitterion (major component)

## Essential Problem-Solving Skills in Chapter 24

1. Correctly name amino acids and peptides, and draw the structures from their names.
2. Use perspective drawings and Fischer projections to show the stereochemistry of D- and L-amino acids.
3. Explain which amino acids are acidic, which are basic, and which are neutral. Use the isoelectric point to predict whether a given amino acid will be positively charged, negatively charged, or neutral at a given pH.
4. Show how one of the following syntheses might be used to make a given amino acid:
   reductive amination
   HVZ followed by ammonia
   Gabriel–malonic ester synthesis
   Strecker synthesis
5. Predict products of the following reactions of amino acids: esterification, acylation, reaction with ninhydrin.
6. Use information from terminal residue analysis and partial hydrolysis to determine the structure of an unknown peptide.
7. Show how solution-phase peptide synthesis or solid-phase peptide synthesis would be used to make a given peptide. Use appropriate protecting groups to prevent unwanted couplings.
8. Discuss and identify the four levels of protein structure (primary, secondary, tertiary, quaternary). Explain how the structure of a protein affects its properties and how denaturation changes the structure.

# Study Problems

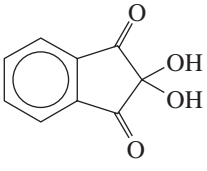
**24-32** Define each term and give an example.

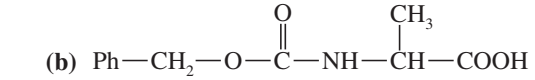
| | | | |
|---|---|---|---|
| **(a)** $\alpha$-amino acid | **(b)** L-amino acid | **(c)** essential amino acid | **(d)** dipolar ion |
| **(e)** isoelectric point | **(f)** Strecker synthesis | **(g)** electrophoresis | **(h)** transamination |
| **(i)** peptide bond | **(j)** hydrogenolysis | **(k)** enzymatic resolution | **(l)** zwitterion |
| **(m)** peptide | **(n)** protein | **(o)** primary structure | **(p)** secondary structure |
| **(q)** tertiary structure | **(r)** quaternary structure | **(s)** pleated sheet | **(t)** $\alpha$ helix |
| **(u)** conjugated protein | **(v)** protein denaturation | **(w)** disulfide bridge | **(x)** Edman degradation |
| **(y)** prosthetic group | **(z)** solid-phase peptide synthesis | **(aa)** oligopeptide | **(bb)** prion protein |

**24-33** Draw the complete structure of the following peptide.

Ser-Gln-Met · $NH_2$

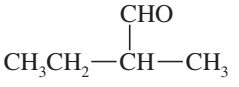
**24-34** Predict the products of the following reactions.

**(a)** Ile +  $\xrightarrow[\text{heat}]{\text{pyridine}}$

**(b)**  $\xrightarrow{H_2, Pd}$

**(c)** Lys + excess $(CH_3CO)_2O \longrightarrow$

**(d)** (D,L)-proline $\xrightarrow[\text{(2) hog kidney acylase, } H_2O]{\text{(1) excess } Ac_2O}$

**(e)**  $\xrightarrow[H_2O]{NH_3, HCN}$

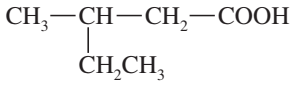
**(f)** product from part (e) $\xrightarrow{H_3O^+}$

**(g)** 4-methylpentanoic acid + $Br_2/PBr_3 \longrightarrow$

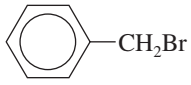
**(h)** product from part (g) + excess $NH_3 \longrightarrow$

**24-35** Show how you would synthesize any of the standard amino acids from each starting material. You may use any necessary reagents.

**(a)** 

**(b)** 

**(c)** 

**(d)** 

**24-36** Show how you would convert alanine to the following derivatives. Show the structure of the product in each case.

**(a)** alanine isopropyl ester **(b)** *N*-benzoylalanine

**(c)** *N*-benzyloxycarbonyl alanine **(d)** *tert*-butyloxycarbonyl alanine

**24-37** Suggest a method for the synthesis of the unnatural D enantiomer of alanine from the readily available L enantiomer of lactic acid.

lactic acid

**24-38** Show how you would use the Gabriel–malonic ester synthesis to make histidine. What stereochemistry would you expect in your synthetic product?

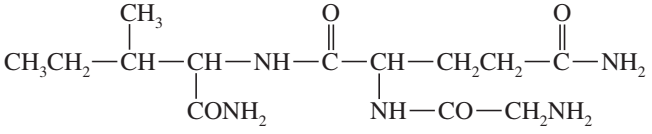
**24-39** Show how you would use the Strecker synthesis to make tryptophan. What stereochemistry would you expect in your synthetic product?

**24-40** Write the complete structures for the following peptides. Tell whether each peptide is acidic, basic, or neutral.

**(a)** methionylthreonine **(b)** threonylmethionine

**(c)** arginylaspartyllysine **(d)** Glu-Cys-Gln

**24-41** The following structure is drawn in an unconventional manner.



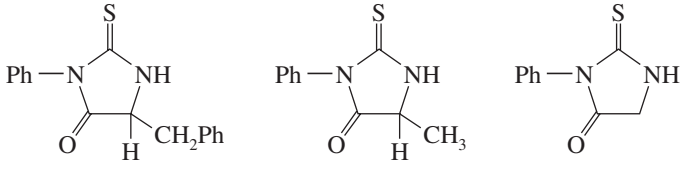
**(a)** Label the N terminus and the C terminus. **(b)** Label the peptide bonds.

**(c)** Identify and label each amino acid present. **(d)** Give the full name and the abbreviated name.

**24-42** *Aspartame* (Nutrasweet®) is a remarkably sweet-tasting dipeptide ester. Complete hydrolysis of aspartame gives phenylalanine, aspartic acid, and methanol. Mild incubation with carboxypeptidase has no effect on aspartame. Treatment of aspartame with phenyl isothiocyanate, followed by mild hydrolysis, gives the phenylthiohydantoin of aspartic acid. Propose a structure for aspartame.

**24-43** A molecular weight determination has shown that an unknown peptide is a pentapeptide, and an amino acid analysis shows that it contains the following residues: one Gly, two Ala, one Met, one Phe. Treatment of the original pentapeptide with carboxypeptidase gives alanine as the first free amino acid released. Sequential treatment of the pentapeptide with phenyl isothiocyanate followed by mild hydrolysis gives the following derivatives:

first time second time third time



Propose a structure for the unknown pentapeptide.

**24-44** Show the steps and intermediates in the synthesis of Leu-Ala-Phe

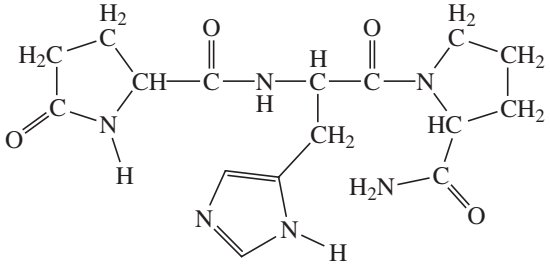
**(a)** by the solution-phase process. **(b)** by the solid-phase process.

**24-45** Using classical solution-phase techniques, show how you would synthesize Ala-Val and then combine it with Ile-Leu-Phe to give Ile-Leu-Phe-Ala-Val.

**24-46** Peptides often have functional groups other than free amino groups at the N terminus and other than carboxyl groups at the C terminus.

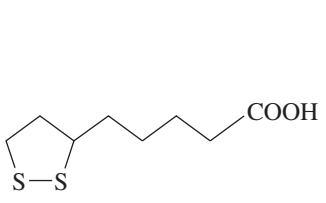
**(a)** A tetrapeptide is hydrolyzed by heating with 6 M HCl, and the hydrolysate is found to contain Ala, Phe, Val, and Glu. When the hydrolysate is neutralized, the odor of ammonia is detected. Explain where this ammonia might have been incorporated in the original peptide.

**(b)** The tripeptide *thyrotropic hormone releasing factor* (TRF) has the full name pyroglutamylhistidylprolinamide. The structure appears here. Explain the functional groups at the N terminus and at the C terminus.

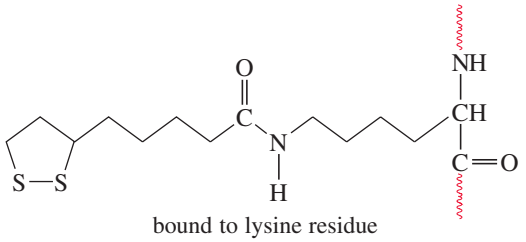


**(c)** On acidic hydrolysis, an unknown pentapeptide gives glycine, alanine, valine, leucine, and isoleucine. No odor of ammonia is detected when the hydrolysate is neutralized. Reaction with phenyl isothiocyanate followed by mild hydrolysis gives *no* phenylthiohydantoin derivative. Incubation with carboxypeptidase has no effect. Explain these findings.

**24-47** Lipoic acid is often found near the active sites of enzymes, usually bound to the peptide by a long, flexible amide linkage with a lysine residue.



lipoic acid

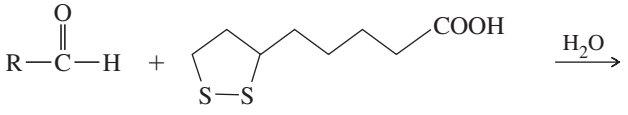


bound to lysine residue

**(a)** Is lipoic acid a mild oxidizing agent or a mild reducing agent? Draw it in both its oxidized and reduced forms.

**(b)** Show how lipoic acid might react with two Cys residues to form a disulfide bridge.

**(c)** Give a balanced equation for the hypothetical oxidation or reduction, as you predicted in part (a), of an aldehyde by lipoic acid.



**24-48** Histidine is an important catalytic residue found at the active sites of many enzymes. In many cases, histidine appears to remove protons or to transfer protons from one location to another.

**(a)** Show which nitrogen atom of the histidine heterocycle is basic and which is not.

**(b)** Use resonance forms to show why the protonated form of histidine is a particularly stable cation.

**(c)** Show the structure that results when histidine accepts a proton on the basic nitrogen of the heterocycle and then is deprotonated on the other heterocyclic nitrogen. Explain how histidine might function as a pipeline to transfer protons between sites within an enzyme and its substrate.

**24-49** Metabolism of arginine produces urea and the rare amino acid *ornithine*. Ornithine has an isoelectric point close to 10. Propose a structure for ornithine.

***24-50** Glutathione (GSH) is a tripeptide that serves as a mild reducing agent to detoxify peroxides and maintain the cysteine residues of hemoglobin and other red blood cell proteins in the reduced state. Complete hydrolysis of glutathione gives Gly, Glu, and Cys. Treatment of glutathione with carboxypeptidase gives glycine as the first free amino acid released. Treatment of glutathione with 2,4-dinitrofluorobenzene (Sanger reagent, page 1178), followed by complete hydrolysis, gives the 2,4-dinitrophenyl derivative of glutamic acid. Treatment of glutathione with phenyl isothiocyanate does not give a recognizable phenylthiohydantoin, however.

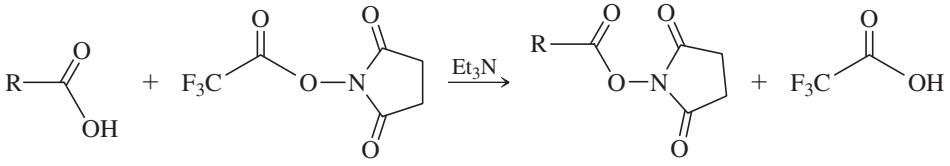
**(a)** Propose a structure for glutathione consistent with this information. Why would glutathione fail to give a normal product from Edman degradation, even though it gives a normal product from the Sanger reagent followed by hydrolysis?
**(b)** Oxidation of glutathione forms glutathione disulfide (GSSG). Propose a structure for glutathione disulfide, and write a balanced equation for the reaction of glutathione with hydrogen peroxide.

**24-51** Complete hydrolysis of an unknown basic decapeptide gives Gly, Ala, Leu, Ile, Phe, Tyr, Glu, Arg, Lys, and Ser. Terminal residue analysis shows that the N terminus is Ala and the C terminus is Ile. Incubation of the decapeptide with chymotrypsin gives two tripeptides, **A** and **B**, and a tetrapeptide, **C**. Amino acid analysis shows that peptide **A** contains Gly, Glu, Tyr, and $NH_3$; peptide **B** contains Ala, Phe, and Lys; and peptide **C** contains Leu, Ile, Ser, and Arg. Terminal residue analysis gives the following results.

| | N terminus | C terminus |
|---|---|---|
| **A** | Gln | Tyr |
| **B** | Ala | Phe |
| **C** | Arg | Ile |

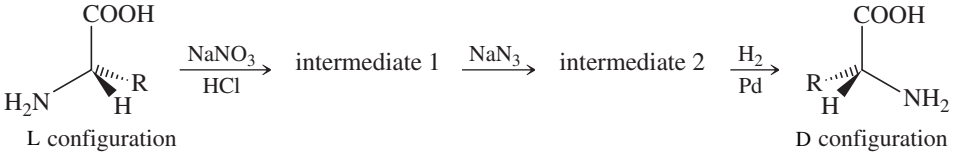
Incubation of the decapeptide with trypsin gives a dipeptide **D**, a pentapeptide **E**, and a tripeptide **F**. Terminal residue analysis of **F** shows that the N terminus is Ser, and the C terminus is Ile. Propose a structure for the decapeptide and for fragments **A** through **F**.

**24-52** There are many methods for activating a carboxylic acid in preparation for coupling with an amine. The following method converts the acid to an *N*-hydroxysuccinimide (NHS) ester.




**(a)** Explain why an NHS ester is much more reactive than a simple alkyl ester.
**(b)** Propose a mechanism for the reaction shown.
**(c)** Propose a mechanism for the reaction of the NHS ester with an amine, $R-NH_2$.

**24-53** Sometimes chemists need the unnatural D enantiomer of an amino acid, often as part of a drug or an insecticide. Most L-amino acids are isolated from proteins, but the D amino acids are rarely found in natural proteins. D-amino acids can be synthesized from the corresponding L amino acids. The following synthetic scheme is one of the possible methods.



L configuration D configuration

**(a)** Draw the structures of intermediates 1 and 2 in this scheme.
**(b)** How do we know that the product is entirely the unnatural D configuration?